

Whose scans are they, anyway?

Raw data are useful for researchers wishing to replicate the results of an experiment. Care needs to be taken when, as with brain-imaging measurements, such data can be misused or misinterpreted.

Like motherhood and apple pie, the concept of sharing primary data is widely recognized among scientists as a good thing. The difficulty lies in putting this laudable aim into practice — and the current controversy surrounding the National fMRI Data Center, a new repository for brain-imaging data at Dartmouth College in Hanover, New Hampshire (see page 445), provides a case in point.

Michael Gazzaniga, director of the centre, is a respected cognitive neuroscientist, and an enthusiast for data sharing. But in his efforts to encourage brain mappers to submit their data to the Dartmouth centre, Gazzaniga has let his enthusiasm run away with him. As editor of the *Journal of Cognitive Neuroscience*, Gazzaniga had the right to announce that his journal would in future require authors of papers containing results obtained using functional magnetic resonance imaging (fMRI) to submit their primary data. But a letter naming other journals as participants in the National fMRI Data Center's mission was a tactical error. It created the erroneous impression that journals were conspiring to force scientists to deposit their data without due consultation, and has alienated a sizeable section of the community that Gazzaniga hopes to serve. Giving the centre a 'national' tag was also provocative — particularly given that many of those working in this area of bioinformatics are annoyed that Gazzaniga and his team did not consult more widely before launching their initiative.

The controversy may have a positive effect, however, if it provokes debate among brain mappers about how best to share data — and what, exactly, should be shared. The Human Genome Project has provided a model of openness, with participating laboratories submitting their DNA sequences to public databases within 24 hours. But what works for one discipline, and one type of data, cannot necessarily be transferred to another. The challenge for brain mappers is to achieve an optimal trade-off between the scientific advances that can accrue from the free exchange of data, and the rights of researchers to benefit

from results they have worked hard to acquire. At the same time, they must devise robust procedures to protect confidentiality, for carelessly archived fMRI data could allow experimental subjects to be identified.

The problem is that fMRI is a young field, in which techniques for data acquisition and analysis are not yet standardized. For instance, some fMRI practitioners argue that data will increasingly be presented as flattened surface representations of the brain. Creating a database with rigid categories that do not allow such representations could in the long term be counterproductive, they argue.

But such difficulties should not be an excuse for delaying discussions on data sharing. Indeed, fMRI's growing pains provide a powerful argument for more openness. The field is prone to artefact, with tiny head movements able to cause spurious signals. With different groups using a variety of analytical techniques to derive their images, it is becoming hard to assess the validity of researchers' claims. And as the technique becomes available to more scientists, many of whom will be less aware of the pitfalls, these problems will grow. Ultimately, data sharing and standardization of tools may be the best solution.

The Organization for Human Brain Mapping (OHBM) has launched a task force to deliberate these issues. This was under discussion before the National fMRI Data Center entered the fray. But Gazzaniga and his colleagues have given the task force's deliberations a new urgency — and a computer scientist from the National fMRI Data Center, Javed Aslam, has been invited to take part.

The OHBM has asked journals to consider the input of its task force before adopting policies on fMRI data. It expects to produce interim recommendations before the Society for Neuroscience meeting in New Orleans in November. *Nature* and *Nature Neuroscience* have no plans to insist that fMRI data relating to their papers are deposited in a public database. But if the OHBM can build a consensus on data sharing, it will have done an emerging discipline a great service. ■

Britain's science bounties

Announcements of enhanced funding for researchers and equipment are almost as good as they look.

Announcements over the past two weeks by the British government of major increases in funding for science (see page 446 and *Nature* 406, 335; 2000) are good news, and go a long way to meet the most urgent needs of the UK science base. Apart from the boost in research output that can now be expected, aspects to celebrate include substantially increased pay for researchers at the top and bottom of their career paths and the significant and more sensibly directed investment into laboratory equipment up to 2004. Since well before his election, Prime Minister Tony Blair has rightly sought a long-term increase in British technological entrepreneurialism, and the new announcement reinforces progress with a boost to training at universities.

People will rightly complain that pay for young scientists needs an even bigger increase, and urgently. A requirement that universities raise at least a quarter of infrastructure finance from non-

government sources is also potentially problematic. Smaller universities may find this too big a challenge. And even the most prestigious ones will need urgently to know just what their targets are if they are to raise external funding over the government's timescale.

One noticeable absence from the announcement is a financial commitment to back up a proclaimed transparency and consultativeness of scientific advisory bodies, on which the government is working over the next few months. But it needs to prepare guidelines of best practice, an implementation plan and a sizeable budget to deliver the significant resources to communication and public consultations that it already knows are required. As the debates over BSE and GM crops showed, there is a huge cost if public confidence is lost. For trusted scientific advice, planning and technology regulation, the British government will need to put taxpayers' money where its mouth is. ■





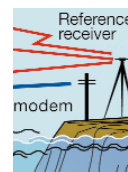
Winged bargain
Fossil of flying reptile goes under the hammer
p446



Algae outbreak
California is wrestling with an algal invasion
p447



Gene green light
Plans for a Berlin genome research centre are approved
p448



Deep views
Future of ocean research is in seafloor observatories
p449

Prospect of data sharing gives brain mappers a headache

Peter Aldhous

An attempt to encourage the sharing of brain images has ignited a fierce controversy about neuroscientists' right to control their data. Some specialists in functional magnetic resonance imaging (fMRI) fear that the National fMRI Data Center, newly established at Dartmouth College in Hanover, New Hampshire, will take over control of their results. Others are concerned that the centre's operation might breach patient confidentiality.

Blood flow and energy consumption in the brain are mapped by fMRI, showing which brain structures are active when particular cognitive tasks are performed. But behind each image lies a mass of raw nuclear magnetic resonance data, statistical analyses and detailed anatomical records.

Having access to these underlying data could help fMRI specialists interpret one another's work. But brain mappers are divided over the merit of sharing primary data, and many of them object to the way in which Michael Gazzaniga, director of the National fMRI Data Center, has set about the task.

In mid-June, Gazzaniga wrote to fMRI specialists who had published in the *Journal of Cognitive Neuroscience*, which he edits. The journal would in future require authors of fMRI papers to submit "all experimental data" to the centre, he said.

The letter named *Science*, the *Journal of Neuroscience*, *Cerebral Cortex* and *Neuron* as participants in the data centre's mission. Although the letter did not say that these journals would adopt a similar policy of compulsion, many recipients feared that this was its implication.

That prompted an anguished letter, coordinated by Isabel Gauthier of Vanderbilt University in Nashville and signed by several dozen fMRI specialists, sent to the data centre's financial backers — the US National Science Foundation and the Keck Foundation — and the editors of 14 leading journals.

These included *Nature* and *Nature Neuroscience*, plus those named in Gazzaniga's letter. "Mandatory submission of fMRI data impinges on the rights authors should have on the publication of findings stemming



Gauthier: wants to keep control.

from their own work," the letter argued. "The nature of fMRI is such that it may be difficult to separate the data that are relevant to a published paper from data that are destined to another manuscript."

"If you have done the work, you should have a chance to benefit from it first," says Gauthier. Her particular concern is for small research groups that lack the resources to conduct all of the analyses they would like before submitting their first paper. Gazzaniga's policy of compulsion, Gauthier claims, would allow other scientists to incorporate the data into their own publications.

Not everyone agrees. "If you have a reason to hide your data, there's a problem," says Roger Tootell of Harvard Medical School.

An archive would allow scientists to spot questionable images more easily, he says.

Nevertheless, by mid-July, the National fMRI Data Center's website had been amended to address Gauthier's letter. Authors would have "the option to place their data in a 'data hold' for a fixed period of time". This would keep the primary data hidden from users of the database. "I think that would be satisfactory to me," says Gauthier, "but I can't tell exactly what would make everyone happy."

But the fMRI community is not yet agreed on which data are worth archiving. "There are many issues about differences in the way data are acquired and analysed," says Jonathan Cohen of Princeton University in New Jersey, a council member of the international Organization for Human Brain Mapping (OHBM).

In an attempt to build a consensus, the OHBM has set up a task force on neuroinformatics. Earlier this week, Cohen wrote to the editors of leading journals asking

Canada selects health institutes

David Spurgeon, Montreal

The Canadian Institutes of Health Research (CIHR) has announced the creation of 13 virtual institutes to link researchers widely separated by geography and discipline. They will provide a long-awaited new structure of support for biomedical research in Canada.

Although the structure of the federation of institutes is bound to invite comparisons with the US National Institutes of Health, Alan Bernstein, president of the CIHR, describes the concept as "uniquely Canadian".

Scientific directors will be guided by an advisory group of Canadians and international experts, and supported by "the best researchers wherever they conduct their work, in Canada or abroad", he says.

The new institutes will be devoted to: aboriginal peoples' health; cancer; circulatory and respiratory systems; gender influences on health; genetics; health services and policy; healthy ageing; human



Healthy response: aboriginal well-being is one of the areas covered by Canada's initiative.

development and child health; infection and immunity; neurosciences, mental health and addiction; musculoskeletal health and arthritis; nutrition, metabolism and diabetes; and population and public health.

► them to consider its deliberations before setting their fMRI data policy.

So far, few have followed the *Journal of Cognitive Neuroscience's* lead. Indeed, none of those named in Gazzaniga's letter have adopted a policy of mandatory submission to the National fMRI Data Center.

Other researchers argue that the issues run deeper than scientists' control of their data. Peter Fox, of the University of Texas Health Science Center in San Antonio, notes that individuals could be identified from the anatomical data requested by Gazzaniga. "It's quite easy to make a 3D model of the face," says Fox, who is working with the US Department of Justice on the potential of fMRI data in forensics.

The National fMRI Data Center's instructions for data submission acknowledge these concerns. "We can provide security measures," says a memo dated 21 June. "We'll be developing algorithms," adds Gazzaniga. But Fox is concerned that the centre has not consulted widely enough. "We put in phone calls, and they weren't returned," he says.

"There was no attempt at exclusion," responds Gazzaniga. "But if there'd been a phone call or two that would have made everyone feel better, we should have done it." Gazzaniga claims that the current controversy has done the field a service by sparking debate on data sharing. "We're provoking some thought," he says. ■

♦ <http://www.fmridc.org>

♦ http://www.psy.vanderbilt.edu/faculty/gauthier/fmridc_letter.html

♦ <http://www.nmr.mgh.harvard.edu/OHBM>

UK hopes that big-money deals will attract top talent

Natasha Loder, London

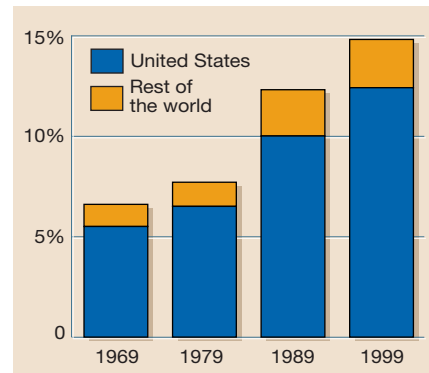
Positions vacant: world-class researchers sought. A combined salary top-up and discretionary grant of up to £100,000 (US\$150,000) available. Apply to Her Majesty's Government.

For years, Britain has worried about both attracting and keeping top scientists. In the 1960s the Labour government even thought about banning foreign job advertisements or treating scientists as an élite group (see *Nature* 403, 121; 2000). The hook has now been baited with money. The idea is that, like sports stars, 50 of the world's leading academics will be tempted to transfer to Britain.

The announcement of the new fund coincides with the release of a white paper in which the government says that Britain needs to attract more scientists and engineers from overseas, and promises to change the immigration rules to make it easier for students and foreign researchers to work and remain in the country.

The salary top-ups are being offered by the UK government, the Wolfson Foundation and the Royal Society — whose fellows have been taking jobs abroad at an increasing rate (see graph). The government will contribute half of an annual £4 million fund.

But the government has offered few details about how the scheme would work, or



Brain drain: a growing proportion of Royal Society fellows work outside Britain.

how long the fund would last. The Royal Society says it intends to make a "long-term commitment". And a panel that includes scientists from the society, the Wolfson Foundation and the Royal Academy of Engineers will be set up to seek out scientists in key areas of research.

Eric Ash, treasurer of the Royal Society, a trustee of the Wolfson Foundation, and one of those behind the scheme, says that the fund would offer money on top of the standard salary offered by the university. "Our feeling is that it won't all be salary. Part of it could be discretionary research grant, which is the hardest thing to get hold of." ■

Science may lose out from sale of 'flying reptile' fossil

Jessa Netting, Washington

The imminent auction of a fossilized gliding reptile that lived 200 million years ago has fuelled concern among palaeontologists



Going, going, gone? The auctioning of *Icarosaurus* may cause it to disappear from researchers' sight.

about the impact of private collectors on their scientific work.

The fossil, *Icarosaurus siefkeri*, was discovered by Alfred Siefker, who, with the help of two teenage friends, dug it out of an abandoned quarry in 1961. It was housed in the American Museum of Natural History in New York until ten years ago, when Siefker reclaimed it and offered it to the museum and other institutions for \$5 million.

The offer was not taken up. Now in ill health, Siefker has dropped the price and opened the bidding to private collectors. The fossil is being auctioned by Butterfields in San Francisco at the end of this month, and is expected to raise \$250,000–\$300,000.

When discovered, the 18-cm-long *Icarosaurus* was the oldest known example of a reptile that had taken to the air, preceding the pterosaurs and birds by as much as 50 million years. Its 25-cm wingspan did not allow powered flight, only gliding.

Palaeontologists have since described an

older gliding reptile, *Weigeltosaurus*.

Despite this, the rarity of 'flying' reptiles, and the fact that Siefker's fossil is the 'type specimen' — the first of its species to be described — make it scientifically valuable.

The prospect of such a fossil falling into private hands has set alarm bells ringing. "The reaction of most people is one of horror," says Mary Dawson, curator of vertebrate palaeontology at the Carnegie Museum of Natural History in Pittsburgh, Pennsylvania. "It would be all right if it goes to a museum, but if not it [could be] lost to science."

David Herskowitz, Butterfields' natural-history specialist, is less pessimistic. "If a person is going to pay enough to put it in their collection, you can be sure when that person dies, it will be donated to a museum," he says. "We are really hoping that they can find a trustee who will pay for it and donate it back to the American Museum in New York." ■

Researchers criticize response to killer algae

Rex Dalton, San Diego

Concerns are being raised about initial efforts to combat the first outbreak on the American Pacific coast of an ecologically damaging marine alga, *Caulerpa taxifolia*, which has already wreaked havoc in the Mediterranean Sea.

Marine biologists are criticizing the absence of the comprehensive research programme that they say is needed to counter the *C. taxifolia* invasion in Carlsbad, on the Californian coast between Los Angeles and San Diego. They want such a programme to be implemented and managed by an independent scientific panel.

They also point to a number of mistakes in the response since the rapidly growing alga — whose fronds displace other species — was discovered in mid-June in a tidal lagoon. One specimen sent to Switzerland for molecular analysis was found to be worthless, and a visiting Dutch researcher was unable to obtain a sample.

C. taxifolia is the latest invasive species to strike at ecosystems around the world, creating fears of ecological and commercial damage. It is a green seaweed with fern-like fronds, and can grow at a daily rate of 8 cm.

Susan Williams, the director of the Bodega Marine Laboratory, run by the University of California (UC) at Davis, says she fears a major opportunity is being lost. She wants to see tests of several control methods, along with more precise data collection.

Although Williams recently joined a technical advisory team of state, federal and private officials overseeing the eradication effort, she claims that her suggestions are being ignored. "I feel like I'm shouting into a hurricane," she says.

Alexandre Meinesz, a biology professor at the University of Nice-Sophia Antipolis and the author of *Killer Algae* (University of Chicago Press, 1999), which describes Europe's experience since the algae struck in 1988, supports Williams' criticisms. He says that the events at Carlsbad sound like "the beginning of my story".

Daniel Simberloff, an ecologist who directs the Institute for Biological Invasions at the University of Tennessee at Knoxville, is also worried about the lack of tests on methods and the delays in molecular analysis of the strain found in Carlsbad.

Greig Peters, a fisheries biologist with the California water quality board, which is leading the eradication effort, says that Williams' "expertise and input has been very valuable", but rejects her criticism: "We have a real cohesive team, responding as rapidly as possible."

Lars Anderson, a plant physiologist from the US Department of Agriculture (USDA), says "an interesting mix" of government offi-



Deadly weed: fronds of *Caulerpa taxifolia* (green) threaten to smother marine ecosystems.

cials and academics has produced tension at a time when "we are not in a research mode".

"But if you stand back and look at other eradication efforts, this one gets high marks," says Anderson, who is playing a leading role in the eradication campaign from his USDA office at a weed-research facility at UC Davis.

There also are signs of a turf war over who fights the algae. Simberloff attempted to have Williams brief a meeting of the federal

Invasive Species Advisory Committee, of which he is a member, in Seattle this week. But the committee of experts advising 20 federal agencies is not expected to invite Williams to talk.

An invitation was seen as possibly offending another federal panel, the Aquatic Nuisance Species Task Force, that met this week in Vermont. Anderson is briefing that group on the Carlsbad experience. ■

France may bid for fusion reactor

Heather McCabe, Paris

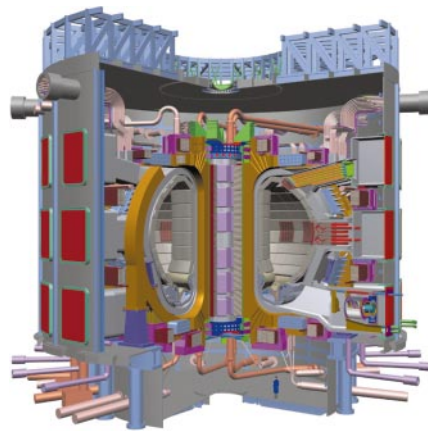
Nuclear scientists and engineers are urging France to compete with Japan and Canada to host ITER (the International Thermonuclear Experimental Reactor), an international collaboration to further fusion research.

At present, there are two candidate sites, one at Clarington in Ontario, Canada, and the other in Japan. Sites in Sweden, Germany and Italy have been discussed, but changing political circumstances mean that none of these are seen as viable.

The French Atomic Energy Commission has now suggested that ITER should be built at its Cadarache site near Aix-en-Provence. But the French government has yet to give the proposal its official backing.

The stumbling block will probably be money; it is sometimes assumed that the host country will pay 50–70% of ITER's estimated cost of at least 3.5 billion euros (US\$3.25 billion). Japanese officials have suggested that their nation would provide as much as two-thirds of the construction costs.

The ITER partners — Europe, Japan and Russia — have welcomed the French proposal. But France may not be favourite to host the reactor, even among Europeans, so the Japanese site has technical advantages. Some fusion



Defused? ITER could stay on the drawing board unless a host nation meets most of its costs.

researchers also hope that a Canadian site would lure the United States, which withdrew in 1999, back into the project.

ITER's next hurdle will be the budget discussions for the European Commission's sixth Framework programme of research at the beginning of next year. The project's supporters will have to convince the European Parliament to back ITER's current design proposal, but the parliament remains deeply divided on fusion energy. ■

Astronomers silenced in star-name wars

William Triplett, Washington

Laurent Pellerin is an angry astronomer. After publicly criticizing one of several organizations that sell individuals the 'right' to name stars, he has been told to shut up by his planetarium employer.

The reason, according to an item in the August issue of *Sky & Telescope* magazine, was that the company — the Illinois-based International Star Registry (ISR) — had threatened Seminole Community College Planetarium in Sanford, Florida, with a defamation lawsuit over Pellerin's remarks, which were broadcast by a local radio station.

Pellerin says he was "only trying to educate the public" to facts that many such companies either play down or do not mention. In particular, when people pay such companies from about \$50 for a certificate conferring the chosen name — often in memory of a dead relative — on a specific star, this has no standing among professional astronomers.

After sending a certificate to the buyer, ISR enters the star's name and coordinates in a book that it publishes. "Because these star names are copyrighted with their telescopic coordinates in the book," ISR's web page claims, "future generations may identify the star name in the directory and, using a telescope, locate the actual star in the sky."

But Bob Milkey, the executive officer of the American Astronomical Society (AAS), says he knows of "no astronomers who look at the ISR catalogue of star names". Despite this, it is not uncommon for visitors to observatories to ask to see 'their' stars.

"One guy came in looking for the star he'd 'named' for his girlfriend," says Robert Martino, assistant director of Perkins Observatory at Ohio Wesleyan University. "I told him there was no star by that name."

The International Astronomical Union is officially the sole designator of stars, typically identified by numbers. When people find that out, Pellerin says, they "can feel as if they've lost their loved one all over again".

Two years ago the New York City Department of Consumer Affairs cited ISR for "engaging in a deceptive trade practice". But ISR has refused to pay the resulting fine of \$3,500 imposed by the city, arguing, according to Rocky Mosele, ISR's vice-president and director of marketing, that "we have never been found in violation of any laws".

ISR has also threatened legal action against at least two other astronomers besides Pellerin over public criticisms. John Martin, a graduate student at Case Western Reserve University, had posted a critical note about ISR on his personal website. "I do a lot of outreach," Martin says, "and I get a lot of questions on this subject. It was meant to be informative."

An ISR lawyer did not see it that way, sending Martin and the university a letter of



objection. Martin deleted the note before the university became involved.

Ohio Wesleyan did get involved — much to Martino's unhappiness. Martino had posted a note on the Perkins Observatory website denouncing star-selling. ISR attorneys threatened to sue not only Martino, but also the observatory and the university if the material were not removed.

Martino claims that the university decided to comply even though it was confident of winning in court. "They gave in because it was cheaper than fighting," he says. But Pam Besel, director of public relations at Ohio Wesleyan University, says administrators ordered Martino to remove the note because "his views are not necessarily the university's".

Pellerin says that senior officials at the Seminole Community College told him not to discuss ISR with anyone. He was later

given a copy of a letter from the university to ISR stating that, although the college's legal department did not believe Pellerin's remarks to be defamatory under Florida law, employees would no longer be talking about ISR. Tom O'Hern, who handles legal matters for the college, declined to comment.

At the AAS, Milkey says he gets "a few" inquiries from people every year asking for advice about buying a star-name. "We're generally very careful in what we say," he notes. "I tell them that it has no standing in the scientific community, that astronomers only recognize IAU designations, but if they want to go ahead and buy it as a novelty, then go ahead."

Others see more at stake. "This is a clear-cut case of academic freedom," says Steve Shore, chairman of the department of physics and astronomy at Indiana University, South Bend.

Mosele is unrepentant. He confirms that ISR has sent letters to various astronomers and institutions threatening suits for defamation/libel. "What they were saying about us was wrong, and it's illegal to say those things," he says. "We were not trying to compromise their rights to free expression; we said to stop saying those things because they're not true."

"In a strict interpretation of the law these companies are not fraudulent," agrees one prominent astronomer who asks not to be identified. "If you pay them money, they'll send you something saying the star is yours; that's what they say they'll do, and they do it. But does the consumer really know what he's buying?" ■

Berlin genome centre to go ahead

Quirin Schiermeier, Munich

After a year of lobbying from local life scientists, the governments of Berlin, Potsdam and the surrounding state have agreed to create the Berlin and Brandenburg Genome Research Centre.

Plans for the centre were drawn up last year by Detlev Ganten, scientific director of the Max Delbrück Centre for Molecular Medicine, Hans Lehrach, director of the Max Planck Institute of Molecular Genetics, Lothar Willmitzer, director of the Max Planck Institute of Molecular Plant Physiology, and Günter Stock, head of research and development at the pharmaceutical company Schering.

Institutes in the region that do genome research will first commit themselves to a joint scientific strategy, linked to the regional biotechnology industry.

© 2000 Macmillan Magazine Ltd. In its three universities, the University of



Ganten: helped draw up plans for the centre.

Potsdam, and other institutes in the region will be overseen by a steering committee consisting of Ganten, Lehrach, Willmitzer, Stock and Volker Erdmann, professor of biochemistry at the Free University of Berlin (FU).

The city and state governments say that the funding of specific projects, particularly in functional genomics and bioinformatics, will then

increase by DM30 million (US\$14 million) per year.

Parts of the 'virtual' genome network will be transferred to a genome centre, probably at the FU, in around 2003. It is designed to house about 150 scientists in 30 cross-institutional research groups. ■

Ocean researchers dive to deep-sea stations

Colin Macilwain, Washington

Fixed seafloor observatories are likely to replace traditional sea-going expeditions aboard research vessels as the dominant source of oceanographic data, leading researchers say. And the US government is starting to make plans for a comprehensive network of such observatories, to give oceanographers as exciting a range of research options as possible.

Oceanographers are increasingly aware of the limitations of their research ship expeditions for data gathering. "They have been very good for things with a very long lead time, such as plate tectonics," says Marcia McNutt, president of the Monterey Bay Aquarium Research Institute and of the American Geophysical Union. "But everything else is characterized by an event, and in expedition mode oceanographers cannot always be in the right place at the right time to collect data. So everyone agrees that observatories are the way to go."

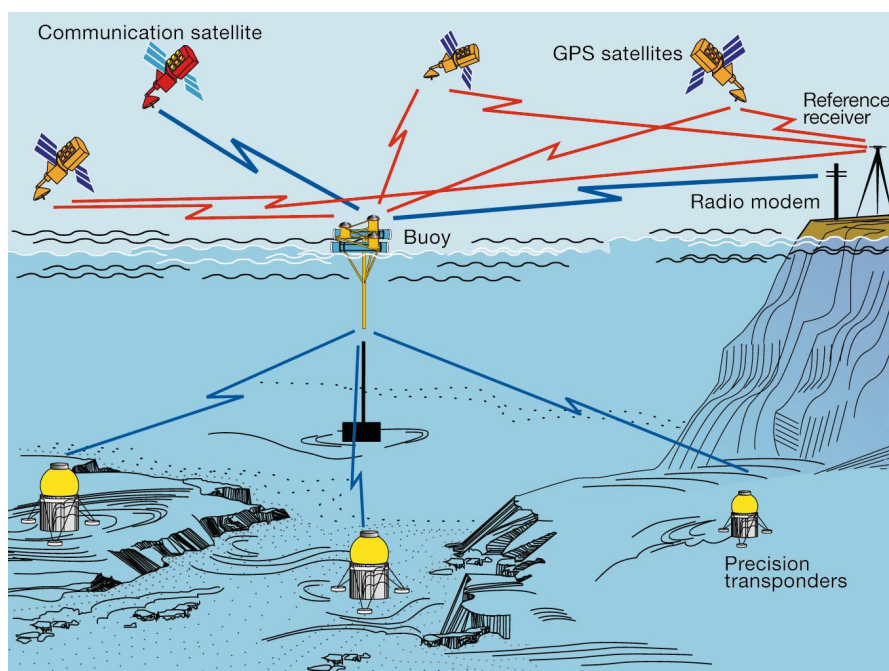
It has been harder, however, to agree on what types of observatories would be most useful, where to put them, who should build and maintain them, and how they can be paid for without overstressing limited research budgets for oceanography.

But late last month, a report to the National Science Foundation (NSF) from the National Research Council (NRC), the operating branch of the National Academy of Sciences, set out a plan of how the problem should be approached. The report, the result of an NRC panel chaired by William Ryan of the Lamont-Doherty Earth Observatory in New York state, proposes that the NSF should plan and build such a system. The recommendation is likely to provide the basis for concrete plans from the NSF and other agencies, which will perhaps be announced next February as part of the 2002 budget.

A first step

The Ryan panel doesn't specify exactly what the network should be, but sets out the main parameters for its construction. "We don't yet have a perfect model, but the NRC report is an important first step," says McNutt, who wasn't on the NRC panel but attended a January workshop in Florida that provided much of the basis for its work.

The report says the network should be flexible and dynamic. It should incorporate some seafloor observatories connected to subsea cables and others with links to fixed buoys on the surface. In addition, rapidly deployable buoys would monitor oceanographic events, from algal blooms to volcanic eruptions. Other suggestions include new grant mechanisms to support very long-term work, open access to all data and an emphasis on standardization, to widen the usefulness



Ocean mission of the future: subsea observatories can send data to shore via buoys and satellites.

of both equipment and data. And it says the NSF should press ahead with its network, without waiting for international agreement on what such a global network should be like.

The panel wasn't asked to look at costs in detail, but it warns that the network will cost "approaching hundreds of millions of dollars" to build and tens of millions of dollars annually to maintain and operate. Oceanographers are confident that the other main agencies in ocean science — the National Oceanic and Atmospheric Administration (NOAA) and the US Navy — will become involved, but say the onus is on the NSF to get the project moving. The NSF's ocean research division is set to use the report to request \$120 million from its Major Research Equipment fund, spread over four or five years agency officials confirm.

CORE, the ocean research lobby group energetically led by Admiral James Watkins, a respected former energy secretary, has helped to build strong support for oceanography in the Congress, where the proposal is likely to be well received. "There's a definite raising of awareness of ocean science," says one congressional staffer.

Most existing seafloor observatories have been built by geophysicists or physical oceanographers. Geophysicists need a lot of communications bandwidth, but not much power for their sensors, and prefer their observatories to be linked by fibre-optic cable to buoys on the surface. Physical oceanographers need more power, which they have obtained by building observatories on subsea cables. But laying new subsea cable over long distances is too expensive for researchers to contemplate.)

sive for researchers to contemplate.)

Researchers interested in the biology and chemistry of the ocean have been slower to use the permanent observatories, and the report says they still lack the sensor technology they need to do this.

Black boxes

But the migration of all of these branches of oceanography to observatory-based data-gathering is inevitable, say the study's authors. Fisheries, pollution and everything else will be observed by little black boxes on the sea floor. "The bottom line is that our ideas and models for explaining phenomena may be stuck if we don't have new methods," says Ryan. "We may be kidding ourselves that intermittent lowering of instruments into the oceans is catching the right things. It is time to look at the oceans in a more rigorous way."

NOAA welcomed the report and expressed its willingness to participate in the network. Jim Baker, its administrator, says the NOAA, NSF and Navy are already discussing ideas for seafloor observatories through an existing forum, the National Oceanographic Partnership Program. He adds that the White House Office of Management and Budget "has left the door open" for an interagency proposal to start building the network next year.

But international collaboration may be a long way off. Several nations, including the United Kingdom and Japan, already have their own networks. "If we want this to happen in anyone's lifetime, the NSF should plan it now and work out international cooperation later," said an NRC study official.

Ecology goes underground

The functioning of terrestrial ecosystems seems to depend heavily on soil biodiversity. But what controls this diversity, and how will it fare in the global greenhouse? Jon Copley digs for some answers.



JEREMY BURGES/SPL

What are the most poorly researched habitats on Earth? Remote environments such as volcanic vents on the ocean floor or lakes hidden under Antarctica's ice must feature prominently on the list. But it might come as a surprise to learn that some ecologists would also include the top metre of dirt in your garden. Terrestrial ecologists have always given at least some thought to the chemistry and physical structure of the soil. But only recently have they begun to understand that soil biodiversity is a crucial factor in regulating how ecosystems function.

Over the past few years, ecologists have realized that many of the most important interactions between plants take place below ground — particularly in the third of the world's soils that are poor in nutrients. In such soils, the dynamic interactions between plant roots, animals and microbial processes seem to determine what grows where and how. But ecologists have barely begun to

study the biodiversity that underpins these interactions. And this ignorance is worrying, as it means nobody can predict how soil biodiversity — and hence the function of terrestrial ecosystems — will be affected by the build-up of greenhouse gases in the atmosphere. Next week, at the Ecological Society of America's annual meeting in Snowbird, Utah, society president Diana Wall of Colorado State University in Fort Collins will call for a major expansion in soil ecology research. "There aren't enough studies out there and they could be multiplied many times over," she says.

As ecologists have begun to pick up their shovels, they have been faced with a huge challenge: a single cubic metre of temperate grassland soil contains thousands of species of microorganisms and invertebrates whose identities and activities are largely unknown. "There is ecology that is as complex as any of the community and ecosystem ecology that has been studied above ground," says David

Tilman of the University of Minnesota in St Paul. Tilman's Minnesota grassland study was instrumental in confirming the link between high biodiversity and both the stability¹ and productivity², in terms of biomass and carbon turnover, of terrestrial ecosystems. But those studies only considered above-ground biodiversity, and Tilman is now turning to the soil in an attempt to understand the real dynamics of his grassland ecosystem.

Turf accounting

Two of the main current projects focusing on soil biodiversity can trace their origins to a summit of British and American ecologists held in 1994 at the Natural History Museum in London, which discussed the need to rectify ecology's subterranean blind spot. In 1998, Britain's Natural Environment Research Council (NERC) started a £6 million (US\$9 million), five-year research programme, centred on a field site

at Sourhope in southern Scotland. Wall, meanwhile, heads a parallel US project, launched in the same year by the National Science Foundation and funded to the tune of \$1.8 million over four years. This is based at a long-term ecological research site at Konza Prairie in Kansas — with some sampling also taking place at Tilman's site at Cedar Creek in Minnesota.

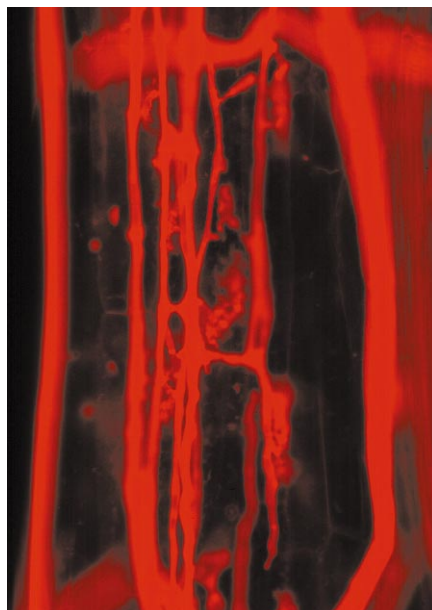
Even before these projects were launched, some ecologists were turning to the tools of molecular biology in an attempt to unlock the secrets of soil diversity. Molecular probes soon revealed that the soil contains many more microorganisms than can readily be cultured³. Now, ecologists are trying to work out which of these microbes is doing what. "The problem with soil microbiology is that you don't know which organisms are doing which job," says Phil Ineson of the University of York, a prominent figure in the NERC soil biodiversity programme.

To tackle this problem, Ineson and his colleagues devised an approach based on the fact that when bacteria copy their DNA before dividing, the new strands of DNA incorporate carbon derived from whatever organic compounds the microbes have been digesting. So by adding particular organic compounds labelled with carbon-13 to bacteria in soil extracts, only those that can break down the compounds should incorporate this heavy isotope into their DNA. Thanks to this weighty DNA, these bacteria can then be separated out by centrifugation, and their DNA sequenced to identify them⁴. "You can characterize the organism that is actually doing the functional business," says Ineson. In May this year, his team demonstrated that a related technique — using gas chromatography and mass spectrometry — could isolate and classify soil bacteria that oxidize atmospheric methane labelled with carbon-13 by detecting the heavy isotope subsequently incorporated into the bacteria's membrane phospholipid fatty acids⁵.

Making the earth move

Other researchers are applying more traditional ecological methods to the problem of soil biodiversity. Experimental manipulations such as removing starfish from a stretch of rocky shore, or excluding sheep from a section of hillside, are relatively easy to do — but controlling the denizens of the soil in a similar way is extremely difficult. "There's no way to radiocollar a bacterium, or to build a cage to keep it out," observes Shahid Naeem of the University of Washington in Seattle, who is examining the microbial ecology of soil from Tilman's plots of Minnesota grassland. "No one can really manipulate bacteria in a natural system, so you usually bring it into the lab."

This is exactly what Hefin Jones of Cardiff University and John Lawton, chief executive of NERC, are doing. Using soil from



Intimate links: the hyphae of an arbuscular mycorrhizal fungus within the root of a plant.

Sourhope, they have set up systems with three different levels of diversity in a facility called the Ecotron. Based at NERC's Centre for Population Biology at Imperial College's outpost in Silwood Park, west of London, the Ecotron is a series of experimental chambers in which conditions can be manipulated according to the whim of the investigator⁶. Jones and Lawton have set up systems at three different levels of diversity, based crudely on the maximum size of the organisms present. Only microorganisms are present in some chambers, whereas others contain microorganisms plus small invertebrates such as springtails and mites. A third set of chambers contain all of the soil biota including larger invertebrates such as earthworms.

The chambers were set up three months ago and the plants will soon be dosed with carbon dioxide labelled with carbon-13 to trace the path of carbon through the different organisms. Mid-way through the 18-month experiment, one half of the microcosms at each level of biodiversity will also be 'shocked' — by adding fertiliser, for example — to see how resilient the systems are to perturbation. Given the rich web of interactions possible between the various soil inhabitants for each treatment, the experiments may well yield more questions than

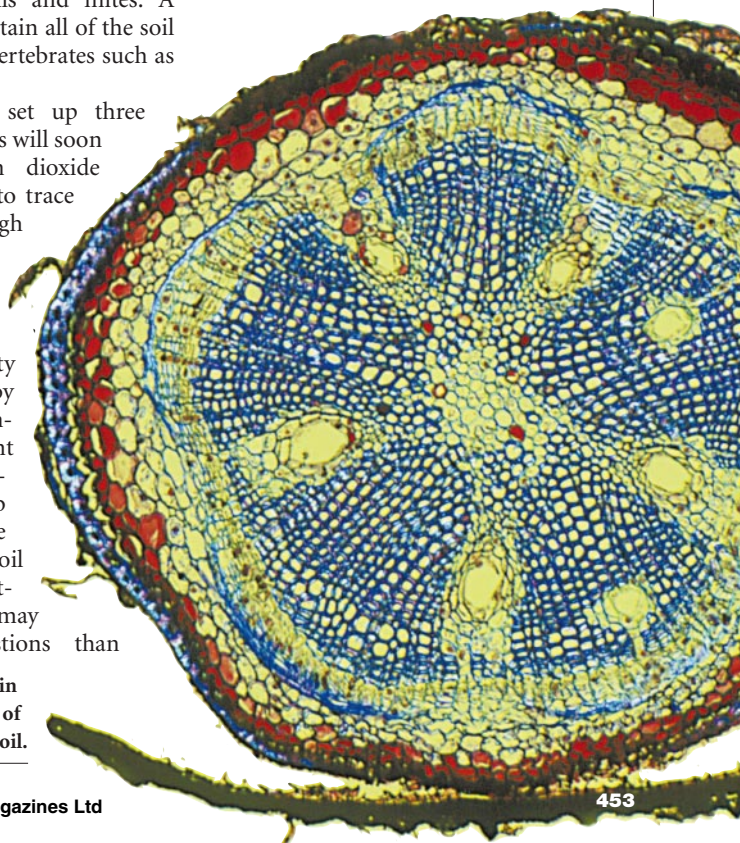
Slice of life: plant roots, seen in section right, are at the heart of dynamic interactions in the soil.

answers. But the results should start to give ecologists a better understanding of the relationship between soil biodiversity and ecosystem function.

'Mycorrhizal' fungi are among the key players in soil ecology. They envelop — and in some cases penetrate — plant roots, and help the plants to acquire hard-to-obtain nutrients such as phosphorus from the soil. In return, the fungi receive carbon from their plant hosts. "If you stand upside down, as it were, and put your head in the soil, you'd think the interesting organism here is the fungus," says Alastair Fitter of the University of York. "There is very good evidence to suggest that the early land plants were only able to grow on land because they were symbiotic with these fungi."

Root of the problem

Mycorrhizal fungi grow through the soil, colonizing the roots of various plants and sometimes forming links between them. In the case of trees colonized by 'ectomycorrhizal' fungi — which surround but do not grow into plant roots — carbon can be transferred from one tree to another through the fungi⁷. 'Arbuscular' mycorrhizae, which actually grow into the roots, forming intimate connections with their cells, influence the diversity and productivity of plants in the soil they inhabit⁸ and protect them from pathogens⁹. They also release a glycoprotein that affects the physical structure of the soil, seeming to boost its stability¹⁰. But the structure and dynamics of these fungal communities are only just beginning to be uncovered. "They are exceptionally important and exceptionally poorly understood,"



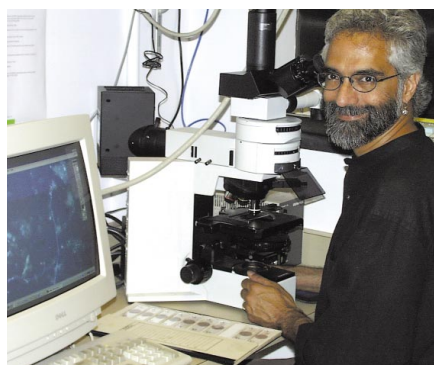
says Fitter. "Because we are above ground, we think from an above-ground perspective. We've forgotten that there is another type of organism down there."

Because they depend on their plant hosts for carbon, mycorrhizae also play an important role in moving the element very rapidly from plants into the soil ecosystem. Fungal filaments, or hyphae, containing carbon obtained from the plant hosts, extend out into the soil and "are immediately going to be gobbled up by every mite, springtail, or whatever there is out there," says Fitter.

To investigate this carbon flow, Ineson and his colleagues have used some 3 kilometres of plastic tubing to pipe carbon dioxide labelled with carbon-13 to plants growing on plots at Sourhope. "We're finding out where the carbon goes, tracking it into mites, earthworms and all the other things," Ineson explains. The researchers have already seen the carbon-13 move swiftly from the plants into the soil¹¹. "Within 30 hours it's gone through the plant," says Ineson. "It goes into the mycorrhizae very quickly and then out into the soil system."

That sinking feeling

Experiments such as this could prove extremely important in providing baseline information from which to investigate how ecosystems will respond to rising levels of greenhouse gases. Soil is an important reservoir for carbon, containing around 1,500 gigatonnes (1,500 × 10⁹ tonnes) globally. At present, it is also a net sink for carbon, absorbing more from the atmosphere than it releases. The big question is what will happen as levels of atmospheric CO₂ rise. How will this affect soil biodiversity? And if this diversity is perturbed, could the soil cease to be a carbon sink? If so, the effects would be dramatic. "There would be another 2 gigatonnes of carbon accumulating in the atmosphere every year," says Fitter. In the 1980s, the amount of CO₂ in the atmosphere was rising by about 3 gigatonnes a year —



There's no way to radiocollar a bacterium, or to build a cage to keep it out — so you usually bring the soil into the lab

Shahid Naeem.

so the collapse of the carbon sink in the soil could potentially wipe out current efforts to reduce greenhouse gas emissions.

In an attempt to answer these questions, Fitter and his colleagues have looked at the effects of elevating levels of atmospheric CO₂, growing turf from a field site on Great Dun Fell in Cumbria, northwest England, in specially designed glasshouses called 'solar domes'¹². "The excitement was below ground," says Fitter. The biomass above ground was unaffected, but when CO₂ levels were raised from 350 to 600 parts per million, the researchers found root biomass increased by up to 50%. The turnover of root growth and death also went up.

Jones, meanwhile, has looked at the effects of elevating CO₂ to similar levels on communities in the Ecotron. As with Fitter's study, the action was underground. "We saw some changes above ground, but the major effects came through in the soil," he says. The species of fungi present changed, with those that digest cellulose becoming more dominant, and total numbers of springtails increased dramatically¹³. "As the animals depend so much on the fungi, changes in the fungi result in changes in soil fauna," Jones explains.

A state of flux

The significance of these changes for carbon flux is not yet clear. Says Jones: "If we believe the soil to be quite a big sink for carbon, do changes in soil biota change the amount of sink?" Fitter agrees that it is unclear whether changes in soil ecology caused by elevated CO₂ will speed up the carbon cycle or slow it down. Higher levels of CO₂ in the atmosphere could lead to a higher ratio of carbon to nitrogen in the soil. Leaf litter with a high carbon-to-nitrogen ratio generally decomposes more slowly, which could slow the release of carbon from the soil, allowing it to accommodate more carbon from the atmosphere over time. If

so, the result could be a form of negative feedback that would put a brake on global warming.

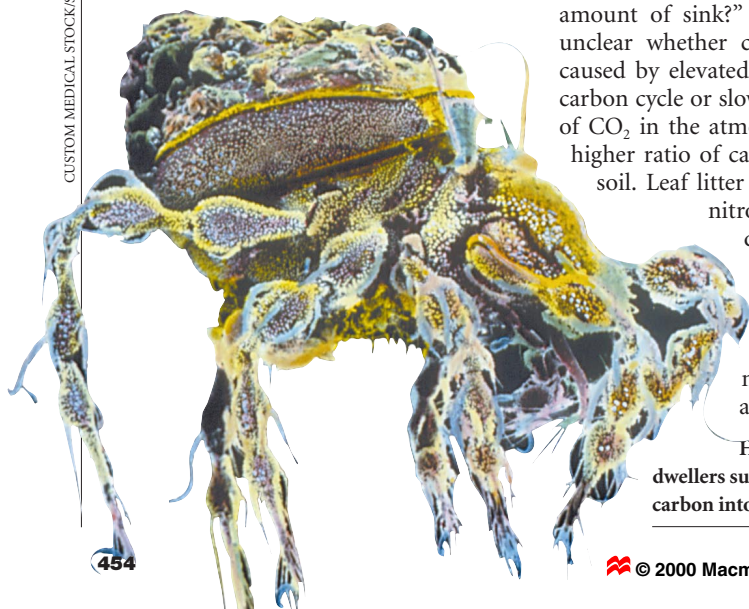
But few ecologists are willing to bet on this happening. "Quite a few experiments suggest that effect, but others show the opposite, so when you put all the data together there is no overall significant effect," says Ineson, who is compiling a review of studies looking at the effects of elevated CO₂ levels on soil carbon flux. "We can't rely on soil to sequester more carbon under elevated CO₂ — this is not the answer to the problem of global warming." Indeed, higher temperatures themselves could cause problems. When Fitter and Ineson experimentally warmed the soil at Great Dun Fell by 3 °C for up to 18 months at a time, they found a hugely increased rate of decomposition¹⁴. This would reduce the soil's ability to act as a store for carbon, potentially exacerbating global warming.

Ecologists working on the problem of soil biodiversity and carbon cycling argue that there is an urgent need for more studies. "We have not yet got the answers to the questions that are going to be so important on a global scale," says Ineson. Without this information, current efforts to limit climate change under the Kyoto Protocol by haggling over land use and trading of 'carbon credits' might be based on shaky foundations. The soil, concludes Ineson, can no longer simply be trampled underfoot. "Suddenly it's coming into the political arena."

Jon Copley is a freelance writer and a teaching fellow at the University of Southampton.

1. Tilman, D. & Downing, J. A. *Nature* **367**, 363–365 (1994).
2. Tilman, D., Wedin, D. & Knops, J. *Nature* **379**, 718–720 (1996).
3. Torsvik, V., Goks, F. J. & Daas, F. L. *Appl. Environ. Microbiol.* **56**, 782–787 (1990).
4. Radajewski, S., Ineson, P., Parekh, N. R. & Murrell, J. C. *Nature* **403**, 646–649 (2000).
5. Bull, I. D., Parekh, N. R., Hall, G. H., Ineson, P. & Evershed, R. P. *Nature* **405**, 175–178 (2000).
6. Lawton, J. H. *et al. Phil. Trans. R. Soc. Lond. B* **341**, 181–194 (1993).
7. Simard, S. W. *et al. Nature* **388**, 579–582 (1997).
8. van der Heijden, M. G. A. *et al. Nature* **396**, 69–72 (1998).
9. Newsham, K. K., Fitter, A. H. & Watkinson, A. R. *J. Ecol.* **83**, 991–1000 (1995).
10. Wright, S. F. & Upadhyaya, A. *Soil Sci.* **161**, 575–586 (1996).
11. Ostle, N., Ineson, P., Benham, D. & Sleep, D. *Rapid Commun. Mass Spectrom.* (in the press).
12. Fitter, A. H. *et al. New Phytol.* **237**, 247–255 (1998).
13. Jones, T. H. *et al. Science* **280**, 441–443 (1998).
14. Fitter, A. H. *et al. Oecologia* **120**, 575–581 (1999).

High turnover: earth dwellers such as this mite move carbon into the soil.



How can the developing world protect itself from biotech patent-holders?

Sir—Biotechnology offers great potential for improving health and food production in the developing world, and achieving both will require significant cooperation between the public and the private sectors.

Such cooperation is being made more difficult by the accumulation of intellectual property rights (IPRs). These include in particular patents on genomic information, and on the basic tools of agricultural biotechnology such as important genes, promoters, and transformation methods. Patent considerations have already had to be taken into account in arranging dissemination of the new vitamin A-enriched rice (see discussions at www.agbioworld.org; archive message 503 gives the details). Not surprisingly, IPRs in biotechnology are highly controversial. Yet there is a dearth of knowledge about their impact, and about appropriate responses.

A concerned group of scientific, legal and economic experts in both agriculture and medicine, from the developed and developing worlds, met at the Rockefeller Foundation's centre at Bellagio in Italy earlier this year to discuss this matter. A number of priorities emerged: the first is to discover whether there really is a problem. Systematic study is needed, for example, of the 'platform' and enabling technologies that are most likely to be valuable to developing nations, to find out whether current IP practices present a barrier to access.

Allegations of direct infringement, such as those made in recent claims lodged by the University of Rochester against pharmaceutical firms marketing Cox-2 inhibitor drugs¹, are not the only concern. We need to know whether the increasing emphasis on IPRs has led researchers to focus on areas likely to maximize royalty income, rather than address developing-world needs.

Information is also needed on whether research institutions, concerned about their relations with donors, are avoiding technologies that they are legally free to use in a limited context. Will international agricultural research institutes, for example, distribute crop varieties containing a *Bt* gene that is unpatented in developing countries, but patented in donor countries?

If there is a problem, the second need is to explore the potential to fine-tune IPR systems. Adjusting standards for granting patents might limit the scope for restricting access to basic or enabling technologies. We need to examine concepts such as 'experimental use' and 'dependency licences', which permit use for certain experimental purposes and for develop follow-on

inventions. These could avert problems such as those faced by the US National Institutes of Health in gaining access to *Cre-lox* technology². The evaluation will include an economic analysis and the implications for health care and agriculture. Important lessons may be learnt from examining past changes in national IP systems.

Other changes also deserve attention. There is an urgent need to examine the legal remedies available to developing countries that find their access to important technologies and products restricted by IPRs—highlighted, for example, by the recent debate over HIV/AIDS drugs in South Africa. Compulsory licences are frequently mentioned in this context. But other approaches exist, such as the public funding of licences, technology exchanges, anti-monopoly mechanisms, and creative use of the leverage available to major donors.

Better understanding of how to manage and license intellectual property in the public sector is also required. When should inventions be licensed exclusively and when non-exclusively? When should inventions originating in the public sector be put in the public domain without legal protection? Some major research organizations such as the International Maize and Wheat Improvement Centre are trying to keep their resources freely available by patenting them before private companies can seize the chance to do so³.

The adoption of the Trade-Related Aspects of Intellectual Property agreement (TRIPs) means the roles of international institutions in area of trade and intellectual property are changing significantly.

The development of international policy needs to be closely studied. This research is likely to lead to proposals for strengthening the participation of developing nations, and modifying procedures for negotiation and resolving disputes.

John H. Barton*, Joseph Strauss†

*George E. Osborne Professor of Law, Stanford University, Stanford, California 94305, USA

†Max Planck Institute for Foreign and International Patent, Copyright and Competition Law, Marstallplatz 1, D-80539 Munich, Germany

1. Malakoff, D. *Science* **288**, 410–411 (2000).

2. Marshall, E. *Science* **281**, 1261 (1998).

3. Dalton, R. *Nature* **404**, 534 (2000).

Cost of institute was small change to Roche

Sir—I find the News report about the sudden closure of the Basel Institute for Immunology (BII) by the Swiss-based drug company Roche (*Nature* **405**, 605; 2000) very sad. Not only is one of the world's leading immunology institutes closing, but also my home town is losing

possibly its best scientific institution.

It is surprising that a giant company such as Roche is not willing to continue to support an institute that has for 30 years helped to establish and maintain the company's scientific reputation. Running the BII costs Fr40 million (\$24 million) a year. This is a small amount compared with the record sales income Roche claimed for 1999: consolidated sales up 12% to Fr27.6 billion and net income up 31% to Fr5.8 billion (see www.roche.com).

It is frustrating that other Swiss academic institutions such as Basel University, Eidgenössische Technische Hochschule or the Swiss National Foundation could not convince Roche to secure an independent existence for the BII through a joint venture.

It is only a partial relief that Roche will replace the BII with a new centre for applied genomics. This will be supervised by Roche's research director, so the institute will lose its independent status. Nobody questions the importance of medically oriented genome research, but closing one of the best immunology institutes in the world is too big a sacrifice. Basel does not have large space resources, but even so Roche could surely have found a site for an additional world-class institute there.

Urs Christen

The Scripps Research Institute, IMM-6, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

How neptunium led to the birth of plutonium

Sir—I read with interest the material from the 13 May 1950 issue of *Nature* reported in "50 Years Ago"¹. However, the statement "In 1940 the first isotopes of the elements 93 (neptunium) and 94 (plutonium) were produced..." is not quite correct.

The uranium sample that ultimately yielded element 94 was bombarded with deuterons in the Berkeley 60-inch cyclotron on 14 December 1940. From the bombarded sample, Glenn Seaborg and co-workers isolated a chemical fraction of element 93 (neptunium) for subsequent studies. To quote Dr Seaborg, "Element 94 was born at last, on the night of February 23–24, 1941."² Actually, element 94 was not named 'plutonium' until March 1942.

Seaborg's interest in element 94 was stimulated by Edwin McMillan's isolation of element 93 in 1940.

C. R. Richmond

Associate Director Emeritus, Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, Tennessee 37830, USA

1. *Nature* **405**, 131 (2000).

2. Seaborg, G. T. "Plutonium Revisited" in *The Radiobiology of Plutonium* (eds Stover, B. J. and Jee, W. S. S.) (J. W. Press, Salt Lake City, 1972).

The road to decarbonized energy

Speeding towards a hydrogen economy — and the obstacles along the way.

Powering the Future: The Ballard Fuel Cell and the Race to Change the World

by Tom Koppel

Wiley: 1999. 276 pp. £16.99, \$27.95

Prospects for Sustainable Energy: A Critical Assessment

by Edward S. Cassedy

Cambridge University Press: 2000. 284 pp. £45, \$69.95

Stanford R. Ovshinsky

Ever since humans first struck flint to ignite fire, energy and technology have been inextricably intertwined. The seeds of science and technology lay in intuitive insight and experiment: these led to new materials that linked energy servo-mechanistically to invention. Yet while materials (and the innovative products made from them) have changed since the Stone, Bronze and Iron ages, the hydrocarbon age is still with us.

The interaction between materials and the products they have spawned through the ages has been greatly affected by the energy source. The burning of coal made railways possible: steam became king and helped usher in the age of electricity. While historians could conclude that electronics transformed the world economy, one could also show that the petroleum-fuelled vehicle industry was the other fundamental base of progress.

Starting with the Industrial Revolution, a basic change has clearly been taking place: energy was being decarbonized from wood to coal to oil and gas, and so on. The curve of energy decarbonization directly overlays that of the growth of modern society. Extrapolating from that curve, it was clear that hydrogen, the ultimate energy source, would not begin its dominance for another 50 years. There are still reserves of oil to be used (although they are depletable), but, as has been pointed out, “the Stone Age did not end because of lack of stone”.

Societal problems jarringly intervene in what appears to be orderly progress. We are familiar with the problems caused by fossil fuels: pollution, climate change, and a strategic dependence on oil that results in global conflict, inflation and the economic burdens of developing countries.

This litany would be an exercise in futility if there were no solutions. But there is a solution that can dramatically shorten the 50-year scenario. Based on hydrogen, it requires a complete systems approach to make it realistic and achievable in the short term. Short term in that we had already entered the hydrogen age some years ago when nickel



Upping the HP and PR: Ford's vice-president Bill Powers extols the virtues of the fuel-cell-powered car.

metal hydride batteries became the enabling technology for electric and hybrid vehicles — electric vehicles can achieve over 200 miles on a single charge and hybrids get 80 miles to the gallon of petrol.

However, there is more than one road to Mecca, because hydrogen can be used as a transition fuel in internal-combustion engines that emit no climate-change gases and in fuel cells that convert chemical into electrical energy. Hydride batteries shuttle ions back and forth and are charged by electricity while fuel cells use up the hydrogen in generating electricity.

Fuel cells have been the centre of attention, but although these are necessary they are not sufficient in the energy equation. Not only must hydrogen be produced in various ways, but it must also be stored and transported safely, economically and practically, without the need for drastic changes in the energy infrastructure. The hydrogen economy must be systems-based with a simple universal means of seamlessly utilizing it. The demonstrably realistic approach that will open a new age is the use of thin-film, multi-junction photovoltaics to obtain hydrogen by breaking up water.

Powering the Future and *Prospects for Sustainable Energy* address the subject from different perspectives. The first is an anecdotal journalistic account of how the proton-exchange membrane (PEM) fuel cell — invented by Sir William Grove in 1839 and developed by General Electric in the 1950s — was redeveloped by a small Canadian team of committed, creative individuals under the inspired leadership of Geoffrey

Ballard. As described in the book, their success required excellent engineering and packaging rather than scientific breakthroughs. It is a gritty, fascinating story of how entrepreneurial and leadership skills combined with engineering and management talents made Ballard Power Systems the leader in fuel cells.

The book has clear limitations, since it accepts without question that the PEM fuel cell is fuel-neutral. It does not put the problems of storage into proper perspective, except to quote an executive: “But think about what would happen here if decent hydrogen storage systems are invented for the car”. Equally important is that the fuel providers for the automotive industry are not likely to provide the methanol infrastructure so blithely discussed, since they can only afford to change their infrastructure once. It is clear to them, and to most of those knowledgeable in fuel cells, that pure hydrogen is required and that this will be very difficult to produce on board the vehicle. The use of gaseous and liquid hydrogen presents problems that limit these two forms to a transient place in the hydrogen economy.

A more feasible approach that is gaining support is the onboard solid storage of hydrogen as a light-weight hydride capable of providing a vehicular range of over 300 miles. This also offers a solution to the important problems of infrastructure, as hydrogen in hydride form is at its lowest free energy and can be transported by conventional means.

Technological progress is based on new materials. The use of hydrogen in a solid is

book reviews

made practical by atomic engineering, as various chemical, electronic and topological functions can be integrated in one material, through such means as catalysis, hydrogen-diffusion paths and acceptor sites.

Edward Cassidy's book is a well-meaning compendium dealing with "the prospects for technological change in the energy industry". It includes a great deal of interesting background information on various forms of energy conversion. Its views are flawed by not being up to date, but it does have value in providing historical reference.

The consequences of the hydrogen economy are of great import, requiring many new books on the subject. Social scientists will seek to understand the dynamics of petroleum companies becoming energy companies: the most progressive of these see themselves as agents of change addressing societal needs. The huge transportation industry will be irrevocably changed as it attempts to respond to societal pressures. New industries based on new science will be developed, and society will benefit. ■

Stanford R. Ovshinsky is at *Energy Conversion Devices*, 1675 West Maple Road, Troy, Michigan 48064, USA.

Romancing the dome

The Private Life of the Brain: Emotions, Consciousness, and the Secret of the Self

by Susan Greenfield

Allen Lane/Wiley: 2000. 272 pp.

£18.99/\$27.95

Rodney Cotterill

There was an awkward moment in Stockholm, in December 1981, when Roger Sperry rose to his feet, intent on rendezvousing mid-stage with King Carl Gustaf. The disease eating away at the ageing neuroscientist's motor function had reduced his gait to a pitifully slow shuffle, and the few-metre journey was going to seem like an eternity. The athletic Swedish monarch — down-to-earth person that he is — defused the situation by bounding over to Sperry to give him his prize. The audience's collectively tensed muscles relaxed. It had been an experience of intense emotion — a subject that provides the core of Susan Greenfield's latest book, which is an attempt to base mind and consciousness on emotions.

Sperry's plight was doubly poignant because he himself had urged a greater focus on the machine's output: the very muscular system that was now failing him. But that advice had been proffered in an article published almost 30 years previously, and had gone largely unheeded. Sperry failed to

Forces to be reckoned with

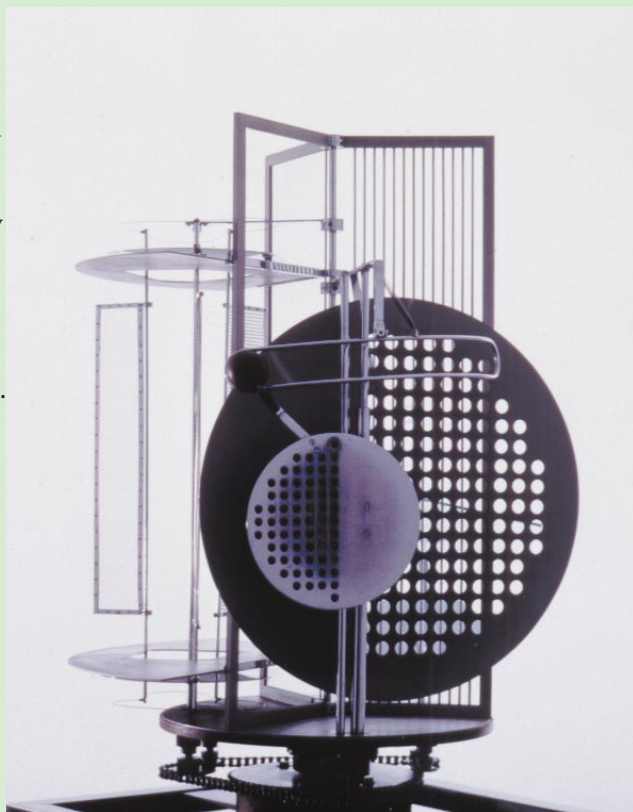
Like distant relations, art and science know about each other but rarely come together. *Force Fields: Phases of the Kinetic*, a new exhibition that opened last month in London, UK,

shows how interesting they can be when they do. The exhibition pays tribute to the effects of twentieth-century physics on twentieth-century art, from the 1920s to the 1980s.

Space, time, light, movement and energy are the dominant themes, represented in mobiles, paintings, installations and film.

There's a touch of the unexpected about much of the exhibition, which makes for interesting viewing (and listening — many of the objects make sounds). Some of the artists are unfamiliar, others better known but represented by unfamiliar works. László Moholy-Nagy's *Light Space Modulator* (shown above), an apparatus for creating an environment of dynamic light projections, is also there.

All the movement and sound help create that rarity, an exhibition that you can take your children to and know they will last more than ten minutes (Jean Tinguely's mobiles have legendary holding power). Another plus is the good representation from Spain and South America, reflecting the exhibition's origins as a collaboration with MACBA, the contemporary art museum in Barcelona. *Force Fields* is at the Hayward Gallery until 17 September (<http://www.hayward.org.uk>). Peter Wrobel



undermine the stimulus-response paradigm: the seemingly common-sense view that sees consciousness as arising in the brain's sensory regions, prior to signals possibly being dispatched to the motor-directing areas. What could be more logical, given that the reader of these words is consciously attending to them but probably not moving? Professor Greenfield belongs to the output-shunning camp, for she emphasizes that her book is about inner emotion, as opposed to mere outward behavioural response. And emotions, she assures us, are the building blocks of consciousness.

Perhaps we should blame Burrhus (B. F.) Skinner for the continuing indifference to motor processes shown by most of those who aspire to solve the mystery of consciousness. Sperry's behavioural wisdom seems to have been thrown out with the behaviourist bathwater. Yet one has only to think of blind deaf-mutes, such as Helen Keller, to appreciate that, for some people, muscular movement is an essential part of observing. The idea that thought itself must implicate an inner surrogate for movement is indeed now gaining ground. This is just as well, for the alternative

so easily leads to *de facto* dualism, with the homunculus hovering in the wings.

The trouble with perceiving movement as a mere product of thought is that it can be misunderstood when it is invoked. "All creatures that move from one place to another have a primitive sort of brain," Greenfield tells us. The bacterium — often-moving, single-celled and thus bereft of neurons — must be glowing with pride. And later we find that the much-vaunted emotions, earlier claimed to be tidily distinct from behavioural response, have become "processes where one is highly interactive with the environment" — apparently as passive freeloaders.

Yet despite these crippling inconsistencies, with their deep implications, the book should succeed in bringing off a difficult double; it should appeal to general readers and also find its way onto some scientists' bookshelves. Popular appeal will never come through megaphoning one's wares from atop the proverbial ivory tower. One has to take one's science out of the laboratory and into the public highways and byways. In Greenfield's case, this has been achieved by,

where appropriate, bravely bringing her private life into the story, by her accessible style and by the wealth of anecdotal matter. Sex occasionally gets into the act, and I found myself wondering about Freudian overtones in the several occurrences of “bump and grind”. Perhaps the old man himself would have relished the prospect of having this book on his couch, although one could imagine his harrumphs over the author’s encroaching on his turf — she listened to the woes of a colleague who had recently broken off an extramarital affair.

The professional appeal will be furthered by Greenfield’s fine documentation, and by her readiness to criticize the ideas of her peers. This is fair enough; if any previous book had solved the big problem, we would all know about it — *ergo*, flawed arguments must abound. Bernard Baars, Derek Parfit, Daniel Dennett, Lawrence Weiskrantz, Gerald Edelman, Marvin Minsky, Igor Aleksander, David Chalmers and William Calvin all come in for polite but determined criticism.

Such balanced evaluation is salutary, so it is bizarre to see almost the entire appendix devoted to the microtubule idea promulgated by Roger Penrose and his leading supporter Stuart Hameroff. The conjecture is that conscious unity stems from long-range quantum coherence in these macaroni-like protein molecules. Faced with this weird story, Greenfield retreats into a timorous “without experimental support”. It is true that she constructively suggests three potential improvements to the model — more coherence, chemical diversity and discrimination between different microtubule types — but she misses the biggest deficit because of her self-imposed limitation discussed above: there’s no coupling to behaviour! Wouldn’t it have been better to have stuck to the common-sense approach that pervades the rest of the book, and warn that an emperor who neglects to clothe himself with even the flimsiest of neuroscientific fact must be in danger of catching a nasty chill?

Given the often subtle way in which emotion has been insinuated into so much of this entertaining book, I have a theory about the schoolboy howler one stumbles over near the start. The story is that of Phineas Gage, probably the most famous patient to survive severe damage to the brain when the force of an explosion propelled a tamping iron through the front of his head. The text has the hapless Gage tamping down dynamite in 1848, 18 years before Nobel invented the stuff. It was gunpowder — dynamite doesn’t need tamping. Could this gem have been slipped in deliberately, just to give nit-picking reviewers an emotional lift? ■

Rodney Cotterill is in the Department of Biophysics, Building 307, Danish Technical University, DK-2800 Lyngby, Denmark.

Trek to an inhabited wilderness

Yellowstone and the Great West: Journals, Letters, and Images from the 1871 Hayden Expedition

edited by Marlene Deahl Merrill
University of Nebraska Press: 2000. 315 pp.
\$29.95, £19.95

Louis Warren

The US government’s surveys of the American West rank among the great scientific endeavours of the nineteenth century. In assessing the American frontier’s potential for commercial development, these surveys were a form of applied science with a government subsidy. Their achievements in geology, palaeontology, cartography and other fields put them among the most influential scientific institutions of the period.

Of all these scientific forays to the West, the 1871 expedition by geologist Ferdinand Hayden to the Yellowstone country has left perhaps the most visible legacy. Hayden’s reports of this and later expeditions were mass-produced and widely read, and did much to bring Yellowstone scenery into middle-class drawing-rooms. Hayden was accompanied by a large retinue, which

included William Henry Jackson, the photographer who first captured much of the Yellowstone country on plates, and Thomas Moran, the artist whose paintings of Yellowstone’s wondrous nature transfixed the public. Together, the geologist, photographer and artist did much to convince US legislators that Yellowstone should be made a national park — the first in the United States — and, subsequently, to persuade Americans to visit it.

And yet, for all the importance of Hayden’s first expedition, we know little about how it operated, or what the party actually did from day to day. Oddly, neither Hayden, Jackson nor Moran kept a diary that survived. It is as if the early nineteenth-century explorers Meriwether Lewis and William Clark had ventured on their mission along the Missouri river to the Pacific and then lost their notes. Marlene Deahl Merrill’s edited version of the journals of two other members of the team, geologist George Allen and mineralogist Albert Peale, thus fills an important gap, and constitutes the first daily account of the historic 1871 Yellowstone Survey. These journals are, in fact, the only diaries of the expedition known to exist. Integrated here with selections from Allen’s field notebook and letters that Peale wrote to newspapers during the expedition, they form a valuable addition to our knowledge of the survey.

This is a good thing, for previous accounts



Nature in hot water

Two hundred feet wide, its adjacent mineral deposits coloured by the microbes that thrive in hot water, the Grand Prismatic Spring in Yellowstone’s Midway Geyser Basin is just one of the 1,000 hydrothermal features to be found

in the national park. *Windows into the Earth: The Geologic Story of Yellowstone and Grand Teton National Parks* by Robert B. Smith and Lee J. Siegel (Oxford University Press, \$27.50) looks into the geological forces responsible.

(most of them written by Hayden and Jackson years later) contain substantial errors. Jackson, for example, recalls in his autobiography that the Hayden party in 1871 was the first group of white men to visit Mammoth Hot Springs, one of the park's central features. How striking, then, to read here that when the party arrived they were met by two white settlers who had claimed the springs and were bent on turning them into a resort for health-seekers — such as the syphilitics who were also there to greet the expedition.

Hayden himself, for a host of reasons, discounted Indian claims to the Yellowstone region, and contributed to the myth of Yellowstone as an “uninhabited wilderness”. In his journal, Peale reports sightings of Indians, and writes of the precautions the party took against Indian attack. (The worry about Indian attack was probably needless, but an idea of what the Indians in question might have looked like is given by a photograph of a family of Bannock Indians taken by Jackson shortly after he left Yellowstone.)

Here, too, are glimpses of the scientific expedition's little-known support staff, people who are invisible in the official reports, including guides, horse wranglers, a driver nicknamed ‘Dummy’ and a Mexican–American hunter known only as José.

This is more than a compilation of journals, however. Interspersed between Allen's sentimental maundering and Peal's scientific enthusiasm are a number of letters from Hayden to Spencer Baird, the assistant secretary of the Smithsonian Institution, reporting on the expedition's progress. Pen-and-ink panoramas of the Yellowstone country, executed by the highly skilled (and self-taught) expedition topographer Henry Elliott grace the headings of each of the 10 chapters, and the work throughout is illustrated with well-chosen photographs by Jackson.

Merrill's editing is superb. In 47 pages of endnotes, she integrates the journals with historical and scientific scholarship about Yellowstone and the American West. The achievements of the expedition become as salient as its curious errors, which include the mismeasuring of various peaks and geological features. The glossary of scientific terms is most helpful to the novice, and the four appendices (which include capsule biographies of expedition members) help flesh out the piece. Those wanting to know about Yellowstone at the time it became a national park have long relied on Hayden's official reports. Now that body of work has an important supplement. Nobody who seeks a deeper understanding of Yellowstone's natural systems as they were in 1871, or how nineteenth-century science was yoked to westward expansion, should miss this remarkable piece of editing and scholarly reconstruction. ■

Louis Warren is in the Department of History, University of California, Davis, California 95616, USA.

Science in culture



Epiphanies of scale

Two versions of the same Einstein statue bear different messages

Haim Watzman

Hidden from the street in a small Jerusalem courtyard, far from the usual routes of tourists and shoppers, is a memorial to Albert Einstein, installed at the end of last year by the Israel National Academy of Sciences and Humanities. Ostensibly it is a scaled-down version of the United States Einstein memorial at the National Academy of Sciences in Washington, not far from the Lincoln Memorial. Yet the Jerusalem Einstein is not the Washington Einstein. Viewing each one is an entirely distinct experience.

The Washington Einstein is a Greek god caught unawares at home. Several times life-size, he dwarfs the visitor. Yes, his hair is tousled and he's wearing a sweatshirt and sandals, but in spirit this statue is a lord of the Universe, gazing down from the equations on the papers in his hand into the heavens at his feet. Like the nearby Lincoln colossus, this Einstein is an object of reverence.

Einstein as a member of the American pantheon could not but be a heterodoxical challenge to Israeli science. For Americans, Einstein was the superhuman genius who came from over the water to establish his abode in Princeton. For Israelis he was a brother, a partner in the founding of their state. He was, as the Israel academy has it in its pamphlet about the monument, “the epitome of twentieth-century Jewish and Zionist aspirations and ideals” — the archetype of the secular, modern Jew who won the admiration of the world but remained wholly and irrevocably a member of his people.

Sculptor Robert Berks, who installed his statue of Einstein in Washington in 1979, says he was well aware of the different context of his Jerusalem work. Now 78 years old, he reworked the statue for Jerusalem in the same way that Titian, Hals and Rembrandt, in their later years, revisited earlier paintings to produce new versions “with more feeling and more data”.

Berks said that when he was asked to duplicate his Washington work he realized that



Monuments to greatness: the Einstein statue in Israel (left) and that in the United States.

the sheer size of the American version was unsuitable for Jerusalem. The buildings surrounding the courtyard are only two storeys high, and there are no buildings more than four storeys high in the immediate vicinity. The Washington giant would have overwhelmed these structures. So he reduced the size by half, creating an Einstein who is still larger than life, but not by much. Einstein in Jerusalem is still awe-inspiring, but you can sit next to him and touch his hand. Children can climb onto his shoulder and try to puzzle out his theoretical scribbles.

The other major difference between the two works is the arrangement of the stars on the black stone map of the sky at his feet. There was an uproar about the map when the statue was being installed in Washington. Berks's original intention had been for the stars to be shown in their position at the moment of Einstein's birth. Prominent scientists signed a petition objecting to the choice of date, arguing that it reeked of astrology. As a result, just a few weeks before the work was unveiled, the map was revised to show the position of the stars on the day of the memorial's dedication.

The Israeli version unashamedly shows the position of the stars at midnight on 14 May 1948, the moment when the State of Israel was established. Einstein has not only been made human — he has been made into the Israeli that, despite his strong Zionist sympathies, he always declined to become.

In Washington, Einstein seems lost in thought on the secrets of the Universe. In Jerusalem he seems to be waiting absent-mindedly for a friend to drop by. In Washington, Berks created Einstein the god; in Jerusalem he created Einstein the Jew — a reminder that physical laws are assumed to be universal, but that images in different contexts may bear different messages. ■

Haim Watzman is a freelance writer in Jerusalem (e-mail: haim@actcom.co.il).

Start making sense

Scientists must stop isolating themselves behind walls of jargon.

Raúl Camba

The past millennium, and in particular the past century, saw revolutionary developments not only in science but also in scientific language. Scientists now speak and write a language alien to most, even to their fellow scientists in other disciplines. Do they realize the ramifications of using a language comprehensible only to the few? Language is a crucial issue for science, especially if scientists want to communicate with the public and other fields. If scientists strive to write simply and clearly, then discoveries will have more impact and attract a wider audience.

Specialist words are essential for accuracy, but they can also result in a private, isolating jargon. For example: "Behavioural effects of menstrual pain are lessened by a common contraceptive" will surely attract more interest than the same paper crippled with the title "Behavioural effects of dysmenorrhoea are attenuated by progestin". I would be willing to bet that only a handful of scientists, let alone lay people, know what dysmenorrhoea is. Yet it appears in the titles of 322 papers referenced in the Science Citation Index. Scientific language overflows with unnecessarily complex terms that alienate all but the most tenacious non-specialist reader. Popular speech, after all, has absorbed only select terms, such as gene, atom, DNA and quantum, while thousands of others — such as dysmenorrhoea — remain obscure to the public.

Simple language, with familiar references, enhances communication. Tycho Brahe's account of the 1572 supernova is exquisite in that it describes the intensity of the stellar explosion relative to the brightness of known celestial bodies: "as bright as Venus" or later "like lead or Saturn". He placed the supernova in the "eighth celestial sphere", and argued against other astronomers who located it at only "15 Earth semi-diameters". In contrast, modern researchers used more complex terminology to describe the 1987 supernova (the pink structure in the centre of the photograph): "a core-collapse supernova which had a peak visual magnitude of 3.0 and was located at 50 kpc from Earth". Tycho and his contemporaries might have lacked better ways to refer to visual intensity or distance, but some of their terms — almost literary flourishes — are still valid today. Modern scientists could have the best of both worlds: precise concepts spiced with well-chosen analogies.

Popular science magazines, an invention of our era, do a good job in spreading the



Compare Tycho's exquisite 1572 account of a supernova with researchers' descriptions in 1987.

word. But more profound changes are required. In many countries, — including mine, Mexico, or the United Kingdom — science students are generally not required to take courses to enhance their written or oral skills. Some scientists are naturally gifted communicators; others, sadly, are not. Given that there is little formal opportunity for the student to better his or her communication skills, the public experience of the maddeningly cryptic scientist is unlikely to change. If scientists are to communicate effectively, then a well-rounded scientific education ought to include courses on how to write an article and give a presentation.

Once they can understand scientists, the public will naturally become more informed and more willing to support initiatives. Furthermore, fewer students will be intimidated by unnecessarily complex language, and cross-fertilization between disciplines will increase. This is the optimistic view; on the other hand, non-anglophone countries face a particularly urgent situation. Once, Latin was the key to the doors of academic institu-

tions. Like it or not, English is now the language of science, and people who don't speak it are effectively cut off from scientific knowledge, except through translations that often appear well after they are most needed.

These countries face a double challenge: educate in science and educate in English. Financial and logistic support from the rich English-speaking countries to spread their language in poorer countries would certainly benefit everyone. Brilliant ideas are frequently wasted if they are produced in the 'wrong' language. Unfortunately, many people still reject English on the grounds of avoiding cultural colonization. But it also has to be said that, in the English-speaking world, the idea persists that anything worth reading must be available in English; and most foreign language journals are regarded as lower class. More openness in both directions is required. Scientists who are conscious of language can do more justice to their disciplines. ■

Raúl Camba is in the Inorganic Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QR, UK.

Taking control

Great moments in megaengineering.

Gregory Benford

2000–2150: Earth stewardship

The runaway greenhouse effect forced Green Puritans to consider planetary management. TwenCen objections to wise human intervention faded as the existing impact of human actions became obvious. Stewardship became essential, following several human diebacks costing more than a billion lives in total.

Changes of cities' reflectivity eliminated the 'heat island' effect, cutting air conditioning costs. The next stewardship stages stimulated cloud generation over tropical oceans, where reflected sunlight efficiency is highest. This went hand in hand with political necessity: economic investment in developing, tropical economies created manufacturing plants that sent particle-rich, cloud-seeding plumes downwind. Capturing carbon dioxide from the air, principally by depositing farm crop waste in the deep oceans, offset the fossil fuel burning of the developing nations. Climate management became routine by 2140, when signs of a coming ice age demanded new measures. Averting this crisis required the spreading of dark soil on polar caps.

Once taken up, managed guidance of a biosphere that supported 12 billion humans could not be renounced without the demise of whole societies. Geospheric stewardship became the greatest moral imperative, and the Puritan Enviro movement was outlawed.

2100–2900: new atmospheres

Similar initiatives on Mars melted the poles, releasing carbon dioxide and water. Discovery of water reserves deep in the martian crust stimulated widespread pumping. Greenhouse build-up in the thin atmosphere further melted some surface deposits, driving a slow accumulation of gases. Funded by the Fogg Foundation, this continued until people could walk the surface wearing only pressure masks.

On the Moon, pioneer drilling uncovered deep ice beds, and exposure of these led to a thin atmosphere. Nanotech agents ferreted out further light elements, releasing them for the first time in 4 billion years. Capping of this atmosphere with a spherical lunar monolayer trapped more, while allowing free transport through gate-gaps.

By this time the most valuable bulk commodity in the inner Solar System had become light elements, gathered from cometary nuclei. Harvesting of volatiles from objects in



the Kuiper Belt and, later, the Oort Cloud, led to the 'Wild West' phase of water prospectors, lighting Earth's night skies with thousands of outgassing tails. Water-rich objects were sold to development firms on Mars and the Moon, flown by their own gas pressures to landing sites, and dropped to create crater-basins. These were domed and outfitted, their soils cultured with nanobacteria, and condominiums opened within 100 days of impact. This proved to be the biggest driver in atmospheric accumulation.

2345: the supernova catastrophe

Early warning that a star 92 light years away was to become a supernova prompted extensive defence measures. To block line-of-sight effects, engineers moved the Moon closer to Earth through artful gravitational flybys of asteroids. While the Farside observatories got a good view of the supernova, projected impact effects upon Earth itself demanded new stewardship measures to offset particle and photon fluxes, and increased cosmic ray heating.

2600: the hydrogen wall

Earth had spent several million years serenely gliding through the low-density Local Bubble, a hole blown by an ancient supernova. Abruptly crossing this edge into a high-density molecular cloud brought a sudden intrusion of interstellar hydrogen into the inner Solar System. Imminent disruption of commerce with the outer, water-rich provinces fuelled a new understanding of both particle and magnetic pressures throughout the heliosphere. This prompted the discovery that the electrodynamic forces of large magnetospheres, principally of the Sun and Jupiter, could be tuned to short-

circuit certain current paths. This brought the flowering of magnetospheric engineering. Closure of cosmic current loops brought protection to the Martian Free States and saved the Asteroid Anarchy. This protection came at the expense of the energy needed to maintain the Jovian discharge, now used to cleanse and mine valuable ices from the Jovian moons, as well as regulating hydrogen intrusion.

2800–present: solar stewardship

The puzzling lack of neutrinos from the solar core was explained when the effects finally percolated to the surface, forcing a re-examination of our understanding of the Sun's evolution. A damping of the interior fires lay ahead, and only intervention could avert a fractional decline in solar emissions. Construction of a plasma torus inside the orbit of Mercury took a century: activation of inductive currents in the torus produced a magnetic field centring on the solar core: cyclic tuning of these fields drove convective circulation between the core and outer shells, refreshing the solar furnace. Transfer of these energies to the surface through magnetic zone effects enabled engineers to overcome the slowing effect of photon diffusion from the core, so that immediate warming of the corona appeared within half a century.

After that, it was a short step to finer tuning, so that emissions could be directed at specific targets in the Solar System. This allowed further warming of Mars and cometary communities which had set up housekeeping in the inner solar system, greatly expanding the human prospect.

3000+: whither the species?

Interstellar travel now promises to extend the human habit of stewardship to the galaxy as a whole. The more radical megaengineers advocate further experiments in the Solar System, but resurgent Puritan Enviro vehemently oppose tinkering with planetary equilibria already struck. In particular, the recent proposal of a 'day without sun' — turning off sunlight as a laboratory experiment, to test new theories — faces widespread opposition. As always, the boundaries of the human reach remain our most vexing moral problem. Yet somehow, the perimeter keeps increasing. Perhaps that is the most enduring lesson. ■

Gregory Benford is professor of physics at the University of California at Irvine. His most recent book is *Deep Time: How Humanity Communicates across Millennia* (Avon).

Biodiversity's ups and downs

Peter J. Morin

Natural ecosystems are devilishly difficult to study. Hence the resort to laboratory simulations, involving short-lived organisms like bacteria, to tackle the question of why some ecosystems support more species than others.

Perhaps the greatest unsolved ecological riddle is why some natural habitats are home to more species than others. Why, for instance, does a hectare of tropical rainforest harbour an estimated 200–300 tree species, while the same area of temperate forest contains only 20–30 tree species (Fig. 1)? One attractive but unproven explanation is that diversity is ultimately determined by the amount of energy available to an ecosystem. On page 508 of this issue, Kassen *et al.*¹ describe clever experiments showing that energy can indeed influence diversity, but that the actual pattern of increasing diversity depends on other environmental factors that favour different species.

Previous support for the influence of energy was indirect, and came largely from diversity–productivity relations compiled for an assortment of ecological communities (here, productivity is a measure of the amount of energy captured and transformed into living matter per unit area and time). These diversity–productivity relations vary at different spatial scales². At large geographical scales (for instance, latitudinally within continents) diversity generally increases with productivity. At smaller local scales (metres to kilometres), such as those within a single abandoned farm field, several different patterns emerge.

Early surveys of the literature³ pointed to a preponderance of unimodal relations between productivity and diversity, especially for terrestrial plants; a unimodal relation is where diversity peaks at intermediate levels of productivity. More recent reviews of a broad range of ecosystems and organisms indicate a variety of patterns, with diversity increasing, decreasing or remaining unchanged⁴ as productivity increases. Although some of these patterns suggest that energy has a causal role⁵, productivity could simply be correlated with other factors that actually generate diversity. One such factor is environmental heterogeneity, which broadly means the spatial or temporal variation in the physical, chemical or biological features of the environment that create different conditions (or niches) that different species can exploit preferentially.

Kassen *et al.*¹ show how diversity can arise and persist even in the simplest systems, and specifically they address the issue of why diversity sometimes peaks at intermediate values of productivity. They used the bac-



Figure 1 Contrasts in species richness. Tropical rainforest (left) contains some ten times more tree species than forests in temperate regions (right). One way to analyse the reasons for this is with purpose-built ecosystems inhabited by short-lived organisms, as Kassen *et al.*¹ have done.

terium *Pseudomonas fluorescens*, which rapidly differentiates into different 'morphs' (here acting as stand-ins for different species) that exploit different microhabitats in unmixed culture vessels⁵. One morph flourishes at the interface between air and liquid growth medium, while another does best in the centre of the culture vessel. Shaking the vessels eliminates environmental heterogeneity and, with it, diversity. When these manipulations are repeated along a gradient of productivity (which can be easily established by varying the concentration of the growth medium), unshaken heterogeneous cultures yield unimodal diversity curves, but these disappear when mixing eliminates environmental heterogeneity.

These observations are consistent with a simple model that predicts that the increase in diversity depends on the niche specialization of different bacterial morphs, and on different rates of production in the different niches. The model is similar to previous ones that predict the coexistence of several species competing for a single resource³, as long as the environment varies and each species

grows best under different environmental conditions.

This explanation is a good deal simpler than the ideas previously floated to explain unimodal diversity–productivity curves. One such idea requires changes in the identity of the limiting resource as productivity increases, together with spatial heterogeneity in resource availability⁶. In terrestrial plants, for example, limitation by soil nutrients at low productivity and by light at high productivity would lead to different low-diversity systems at either end of the gradient, dominated respectively by the best competitors for nutrients or light. However, intermediate levels of productivity combined with resource heterogeneity might allow various species to coexist.

A second idea requires interactions with predators (herbivores as well as carnivores), which increase in intensity as productivity increases^{7,8}. Predators can maintain diversity among their prey by reducing interspecific competition at intermediate levels of productivity. This mechanism breaks down in low- and high-productivity environments,

where predators are respectively too infrequent to thin their prey or so numerous that only the best-defended prey persist. Although the study of Kassen *et al.*¹ does not exclude the possibility that such mechanisms operate in other systems, it shows that neither complication is required to generate unimodal diversity–productivity curves.

Having established the minimal conditions required to produce such curves, challenges remain. It is not clear whether a single mechanistic framework can readily explain other kinds of diversity–productivity patterns. The cause of diversity patterns at larger spatial scales also remains frustratingly uncertain².

In particular, an especially contentious debate continues over the extent of feedback between diversity and productivity. Productivity (measured as energy or nutrient availability) can clearly influence diversity. But differences in diversity among habitats receiving similar amounts of energy can also influence productivity, when productivity is measured as the biomass of organisms accumulating on those sites^{9,10}. One possible feedback mechanism is that increased productivity produces a physical scaffold to support biological heterogeneity (as, for example, in the spatial complexity of forest canopies) on which other species can build. Kassen *et al.* provide a simple example of this kind of self-organization by showing that self-supporting surface films of one bacterial morph form only at higher productivities.

Spectacularly complex natural ecosystems, such as tropical rainforests and coral reefs, have alternately inspired speculations about the causes of diversity and stymied the kinds of experiments required to sort pattern from process. Long-lived organisms, such as trees and corals, simply respond too slowly to experimental manipulations of productivity; it would take many human generations for responses to become apparent, even assuming that alterations of productivity at the appropriate scale could be engineered. That is why purpose-built ecosystems composed of short-lived organisms, such as those used by Kassen *et al.*, are playing an increasingly prominent part in tests of ecological theory. ■

Peter J. Morin is in the Department of Ecology, Evolution, and Natural Resources, 14 College Farm Road, Cook College, Rutgers University, New Brunswick, New Jersey 08901, USA.

e-mail: pjmorin@rci.rutgers.edu

1. Kassen, R., Buckling, A., Bell, G. & Rainey, P. B. *Nature* **406**, 508–512 (2000).
2. Gaston, K. J. *Nature* **405**, 220–227 (2000).
3. Tilman, D. & Pacala, S. in *Species Diversity in Ecological Communities* (eds Ricklefs, R. E. & Schluter, D.) 13–25 (Univ. Chicago Press, 1993).
4. Waide, R. B. *et al. Annu. Rev. Ecol. Syst.* **30**, 257–300 (1999).
5. Rainey, P. B. & Travisano, M. *Nature* **394**, 69–72 (1998).
6. Tilman, D. *Resource Competition and Community Structure* (Princeton Univ. Press, 1982).
7. Holt, R. D., Grover, J. & Tilman, D. *Am. Nat.* **144**, 741–771 (1994).
8. Leibold, M. A. *Am. Nat.* **147**, 784–812 (1996).
9. Tilman, D., Wedin, D. & Knops, J. *Nature* **379**, 718–720 (1996).
10. Hector, A. *et al. Science* **286**, 1123–1127 (1999).

anaemia and the formation of cataracts in the eye.

Although the first small crystallites involved in nucleation form in the bulk of the solution, Yau and Vekilov can observe them only once they have fallen to the bottom of the vessel and attached themselves to the surface there. How can the authors be sure that they are observing the critical early stages of crystallization described by nucleation? A small crystallite will tend to redissolve because it has a large surface to volume ratio; a large enough one will tend to grow in a supersaturated solution because of the thermodynamic driving force towards crystallization. The ‘critical nucleus’ is the size in between, where the probability of adding a particle exactly balances the probability of losing one.

Yau and Vekilov use atomic force microscopy to examine what happens to individual crystallites at the surface. They find that some (the smaller crystallites) tend to lose particles more readily while others (the larger ones) tend to grow. These measurements give a range for the size of the critical nucleus of between 20 and 50 molecules, depending on the concentration of protein in solution. Up to this point, their results simply confirm the long-held assumptions of nucleation theory, and provide some of the first direct observations of critical nuclei.

There is a surprise in the new results, however. For relatively spherical molecules, such as apoferritin, the assumption has always been that the critical nucleus would also be roughly spherical¹: a tiny ball cut out from the bulk crystal lattice. Instead, the authors find that the critical nucleus is more like a raft, consisting of a nearly planar layer of crystal with a partial second layer on top. The shape of the critical nucleus could have a large effect on its energy and therefore on the rate of formation of crystals. If critical nuclei are not spherical, classical nucleation theory does not apply and simple prediction of nucleation rates will no longer be possible.

Yau and Vekilov build on this qualitative picture of crystallization by measuring the nucleation rates of their crystals, and show that their data are consistent with earlier light-scattering measurements of apoferritin nucleation³. It will be interesting to see whether these techniques can be used to determine the effect of protein concentration (and, hence, the degree of supersaturation) on nucleation rates. This would allow a direct test of the nucleation theorem⁴, which is a fundamental relationship used by experimentalists to determine the size of a critical nucleus without actually seeing it.

So classical nucleation theory suggests that the critical nucleus is spherical, whereas Yau and Vekilov find a nearly planar crystalline structure. A third possibility is that the critical nucleus is not crystalline at all, but rather a disordered, liquid-like aggregate

Phase transitions

Catching crystals at birth

David W. Oxtoby

Seeing a crystal when it first appears in solution is no easy task. By the time the crystal is visible, it has already grown to macroscopic size. Using a microscope is not helpful either, because the chance that the crystal will first appear in the volume being examined is very small. Yet the earliest stages of crystallization — referred to as nucleation — usually determine the rate at which crystallization takes place, and therefore the eventual size and purity of the crystals. Since the 1930s, scientists have been developing increasingly sophisticated theories of nucleation, and have measured rates of nucleation experimentally by counting crystals once they have grown to macroscopic size, without ever actually seeing the nucleation events themselves.

On page 494 of this issue, Yau and Vekilov¹ report direct observations of small (100-nanometre scale) ‘crystallites’ taking part in these very first stages of crystallization.

The authors look in detail at the crystallization of the protein apoferritin, which has a relatively slow timescale for crystal growth because of its large size. An ordinary ionic solid will crystallize very rapidly from solution once the first seed appears, because ions diffuse rapidly through water, but protein molecules move slowly enough that small crystallites can be caught before they have grown to macroscopic dimensions. Moreover, the molecules are large enough (a few nanometres) for their individual positions in the crystallite to be determined by modern atomic force microscopy. Proteins are thus useful models for studying crystallization, and they are also of interest in their own right — the crystallization of proteins is the key to determining protein structure by X-ray diffraction, and yet crystallization is still, in many ways, more an art than a science. Protein crystallization is also implicated in certain diseases such as sickle-cell

of molecules. This last structure was predicted for the crystallization of proteins and other large molecules by computer simulations⁵ and the approximate statistical-mechanical method of density functional theory⁶. These studies showed that under certain conditions the nucleation rate was significantly increased (by as much as 30 orders of magnitude). The key to the higher nucleation rates was the existence of a 'metastable' critical point at which fluctuations in protein concentration were enhanced. Under these conditions, it was predicted that the resulting critical nuclei would be highly disordered, with crystallinity appearing only at later stages in the growth

process. It will be revealing to see whether this mechanism for crystal nucleation in proteins can be observed using the methods of Yau and Vekilov, by tuning the experimental conditions to bring their solutions close to a metastable critical point.

David W. Oxtoby is in the James Franck Institute and the Department of Chemistry, University of Chicago, Chicago, Illinois 60637, USA.
e-mail: d-oxtoby@uchicago.edu

1. Yau, S.-T. & Vekilov, P. G. *Nature* **406**, 494–497 (2000).
2. Turnbull, D. J. *Chem. Phys.* **17**, 71–73 (1949).
3. Malkin, A. J. & McPherson, A. *Acta Crystallogr. D* **50**, 385–395 (1994).
4. Kashchiev, D. J. *Chem. Phys.* **76**, 5098–5102 (1982).
5. ten Wolde, P. R. & Frenkel, D. *Science* **277**, 1975–1978 (1997).
6. Talanquer, V. & Oxtoby, D. W. *J. Chem. Phys.* **109**, 223–227 (1998).

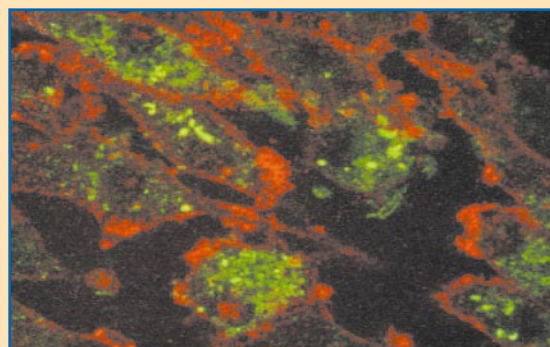
Alzheimer's disease

Plaque removers and shakers

Last year, Dale Schenk and colleagues reported (*Nature* **400**, 173–177; 1999) that Alzheimer's-like pathology in mice could be halted by vaccination with amyloid- β peptide — the main constituent of the amyloid plaques typically found in brains of Alzheimer's patients. Breathtakingly simple, but how does it work? Writing in *Nature Medicine* (**6**, 916–919; 2000), the same group now reveals that antibodies 'decorate' plaques and trigger their clearance by so-called microglial cells in the brain.

Alzheimer's disease is caused by the progressive malfunction and death of nerve cells. The gene encoding amyloid precursor protein (APP) is one of four genes with firm links to the development of Alzheimer's disease, and is mutated in certain hereditary forms of the disorder. How mutations in the APP gene result in the death of nerve cells is still a mystery, but snippets of APP, typically 42 amino acids long ($A\beta_{1-42}$), accumulate within plaques in the brains of Alzheimer's patients, and disease-linked mutations result in overproduction of this APP fragment.

Mice that carry a mutant human APP gene also develop plaques as they grow older — just like humans. Previously, Schenk *et al.* injected such mice with human $A\beta_{1-42}$, and



found that the animals produced antibodies against the peptide, to great effect: plaques were largely prevented from forming, and some of the pre-existing plaques in older mice were even dissolved. The same group now reports that antibodies that recognize $A\beta_{1-42}$, injected into the peritoneum, work as well as immunization with the $A\beta_{1-42}$ peptide. Apparently, the antibodies are able to penetrate the brain in low but therapeutically meaningful amounts.

The authors also show that the antibodies effective *in vivo* bind to amyloid plaques in brain tissue sections of afflicted mice and men, and to antibody receptors on microglial cells. Once activated in this way, these cells engulf and destroy the $A\beta_{1-42}$ peptide, as pictured here ($A\beta_{1-42}$ is shown in green; microglial cells are stained red).

Last week the same group announced initial results from

clinical trials showing that the $A\beta_{1-42}$ vaccine is safe and well tolerated in humans. The results are promising, but these are early days and caution is warranted. First, it's not certain that humans mount as vigorous an immune response to the human $A\beta_{1-42}$ peptide as do mice. If they do not, passive immunization with antibodies may be the way forward.

Second, the mice used for these studies only partially mimic the human disease: although $A\beta_{1-42}$ accumulates in their brains, mice show neither the same loss of nerve cells nor the behavioural abnormalities associated with the human condition. And it's not yet clear whether the amyloid plaques are causing neuronal malfunction and death, or are merely a by-product of the disease, so it remains to be seen whether preventing the plaques from forming will be beneficial to Alzheimer's patients. **Marie-Thérèse Heemels**



100 YEARS AGO

Captain R. H. Elliott, who has been for some time conducting researches into the nature and action of snake venom in India, arrives at the following conclusions in the *British Medical Journal*: — (1) The snakemen of South India are certainly ignorant of any method of producing in themselves a highly-developed condition of immunity. (2) Some few of them appear to practise the swallowing of venom, or the inunction of venom into their limbs, but it is doubtful if they do so with any well-defined object. It is possible that they thus obtain some degree of immunisation. (3) They confine themselves almost exclusively to the cobra, and escape harm by their intimate knowledge of the methods of handling this snake.
From *Nature* 2 August 1900.

50 YEARS AGO

There is anxiety throughout the world concerning reserves of energy. The demand for electricity has increased of late years at an exponential rate, and if the demand for coal, oil and gas has more nearly followed a straight-line law, the slope of the line has been such as to cause concern among individual nations as to when their own supplies of fossil fuels will become exhausted, and to the world in general as to possible sources of energy when there is no more coal or oil... In consequence of this position, there is the greatest activity all over the world to eke out coal reserves by using other sources of energy. Many nations, among whom are the Norwegians, the Swedes, the Canadians, the Finns, the Poles, the Austrians, the Portuguese, the French, the Swiss, and even the Americans, are developing their indigenous sources of water-power; in many of these lands this meets most of the energy requirements. Other ideas are finding practical expression. There is a project at Abidjan on the Ivory Coast for using the potential thermal energy of sea-water existing by reason of the vertical temperature gradient of 68° F. or more over a depth of 1,640 ft. in tropical seas; a boiler heated by water at a temperature of 82° F. would supply a turbine with steam at a corresponding pressure exhausting into a condenser cooled by water at 46° F. A unique hydro-electric scheme is proposed in North Africa whereby the waters of the Qattara Depression, providing some 300 MW. of hydro-electric capacity.
From *Nature* 5 August 1950.

Cancer

Molecular switches in metastasis

Anne Ridley

For patients with solid tumours, the biggest threat to survival is metastasis — the spread of tumour cells from the original growth to other sites in the body. For biologists studying cancer, a major challenge is to identify the underlying molecular changes that switch cells to a metastatic state, with the ultimate aim being to devise treatments that inhibit metastasis. Previous research has concentrated on the contribution of individual genes to metastasis. Now, gene-expression profiling, using high-density DNA microarrays, is revolutionizing our approach to studying cancer.

On page 532 of this issue, Clark *et al.*¹ describe how they used this approach to identify several genes that are selectively upregulated in metastatic mouse and human melanoma cells compared with their non-metastatic counterparts. They find that, remarkably, overexpression of one of these genes — *RhoC* — can stimulate metastasis all by itself. Meanwhile, Bittner *et al.*² (page 536 of this issue) have used microarrays to compare different subgroups of human melanoma, and also find a distinct pattern of gene expression in highly invasive melanoma cells.

In DNA microarrays, probes for the messenger RNA products of up to 10,000 different genes are present on a single 'chip', usually a glass slide. The chips are used to determine which of these genes are expressed (that is, which are transcribed into mRNA) in a selected cell type³. This technology has already been used to classify cancers, such as leukaemia, according to their gene-expression profile^{3,4}.

As their starting material, Clark *et al.*³ used two poorly metastatic cell lines derived from a human and a mouse melanoma. Using an *in vivo* selection procedure in mice (see Fig. 1 on page 532), they isolated highly metastatic variants of each cell line, and compared the expression profiles of 7,070 human genes or 6,347 mouse genes in three different human or mouse metastases with their non-metastatic counterparts. Of course, these genes represent just a fraction of those in the respective genomes. But only 16 of the human genes studied — and a similar number of the mouse genes — were expressed at a significantly higher level in the metastases than in the primary tumours. This small number indicates that the products of these genes are likely to be important in metastasis, rather than being upregulated as an indirect consequence of the altered cell phenotype.

Most of these upregulated genes were not found in both the human and the mouse metastases. Some of the genes, however, were

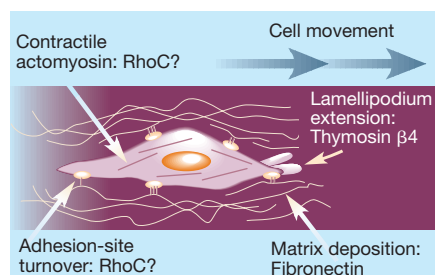


Figure 1 The results of Clark *et al.*¹ and Bittner *et al.*² indicate that the proteins *RhoC*¹, fibronectin^{1,2} and thymosin $\beta 4$ (ref. 1) are required for the metastasis, or spread, of melanoma cells. Cell migration involves protrusion of the plasma membrane (lamellipodium extension) at the leading edge of the cell; the formation of new sites of adhesion to the extracellular matrix at the front; the release of old adhesion sites at the back; and contraction of actomyosin-based cytoskeletal filaments in the cell body to move the bulk of the cell forward. Rho proteins regulate actomyosin-based contractility and adhesion turnover, and *RhoC* might be involved in enhancing these steps of cell migration. Thymosin $\beta 4$ buffers monomeric actin in cells, and could act to provide actin monomers for rapid polymerization into actomyosin filaments in lamellipodia. Fibronectin is an extracellular matrix protein, and deposition of fibronectin by the cell could promote migration by signalling through specific receptors on the cell surface.

not actually present on both arrays. It is perhaps not surprising that there is no unique gene-expression 'fingerprint' for metastatic melanomas: Bittner *et al.*² show that the gene-expression profiles of different subgroups of human melanoma vary greatly. What is significant is that three of the genes identified by Clark *et al.* — those encoding fibronectin, thymosin $\beta 4$ and *RhoC* — showed increased expression in all of the human and mouse melanoma-derived metastases¹.

What do these three proteins do, and why might they be upregulated in metastases (Fig. 1)? Fibronectin is a component of the extracellular matrix and promotes the migration of several cell types, including melanoma cells⁵. The production of fibronectin by melanoma cells might allow them to lay down their own promigratory matrix. Interestingly, increased fibronectin expression also correlated with higher invasive capacity in the study by Bittner *et al.*² (which did not, however, include *RhoC* or thymosin $\beta 4$).

Thymosin $\beta 4$ binds monomeric actin, a component of the cytoskeleton, and may act as an actin buffer, preventing spontaneous

polymerization of actin monomers into filaments but supplying a pool of actin monomer when the cell needs filaments⁶. It is not clear how increased expression of thymosin $\beta 4$ might promote metastasis, but it is likely to relate to the need for cells to migrate.

Clark *et al.*¹ chose *RhoC* for further study. *RhoC* is a member of the Rho family of small GTP-hydrolysing proteins, several of which are known to regulate cell migration^{7,8}; *RhoC* is highly related to *RhoA* and *RhoB*. Unlike the related *Ras* genes, the three *Rho* genes have not yet been found to be mutated in human cancers, but they can contribute to cell transformation to a cancerous state, motility and invasion, at least under experimental conditions^{8,9}. However, the expression of *RhoA* and *RhoB* is not altered in metastatic melanoma cells¹, and it will be interesting to find out how *RhoC* expression is specifically increased.

Clark *et al.* also find that simply overexpressing *RhoC* *in vitro* in the poorly metastatic human melanoma cell line can induce it to become highly metastatic. This is surprising because several other genes are also upregulated consistently in the metastases, and one would expect these genes to contribute to metastasis. Perhaps the higher level of overexpression of *RhoC* in the transfected cells compared with the metastases allows *RhoC* to induce metastasis in the absence of other changes. The poorly metastatic cells are on the verge of metastasis, and it probably takes little to push them further. It would be interesting to know the gene-expression profile of the transfected cells. Is overexpression of *RhoC* sufficient to induce the increased expression of the other metastasis-associated genes, such as those encoding fibronectin or thymosin $\beta 4$? Or is the pattern of expression now different?

It is perhaps unexpected that increased expression of *RhoC* enhances the *in vitro* migration and invasion of melanoma cells, given that constitutively active *RhoA* inhibits the migration or invasion of several cell types^{10–12}. However, *RhoA* is not as potent at inducing melanoma metastasis or migration as *RhoC*¹, and the effect of *RhoC* on cell migration *in vitro* has not previously been studied. These results point to a subtle difference in the properties of *RhoA* and *RhoC*. Also, it is wild-type, not constitutively active, *RhoC* that is overexpressed in the transfected melanoma cells. Wild-type Rho proteins cycle between a GTP-bound and a GDP-bound conformation, whereas constitutively active mutants are locked in the GTP-bound form. Cycling may be important for Rho proteins to promote cell migration. Indeed, a 'fast-cycling' mutant of *RhoA* has considerably greater cell-transforming potential in fibroblasts than does constitutively active *RhoA*¹³.

One drawback to microarray analysis is that it cannot identify post-transcriptional

changes in protein expression or activity. Mutant Ras, for example, can contribute to cancer but would not be identified by microarray analysis of human tumours, as only its activity — not its expression level — is altered by mutation. But the approach taken by Clark *et al.* and by Bittner *et al.* does provide an unbiased method by which to pinpoint important, and potentially new, contributors to cancer. It should now be possible to work back from these transcriptional changes to identify the one or more key genetic or epigenetic events that induce metastasis. The challenge will be to extend this work from a mouse model of metastasis to human patients. ■

Anne Ridley is at the Ludwig Institute for Cancer

Research, University College London, 91 Riding House Street, London W1W 7BS, UK.

e-mail: anne@ludwig.ucl.ac.uk

1. Clark, E. A., Golub, T. R., Lander, E. S. & Hynes, R. O. *Nature* **406**, 532–535 (2000).
2. Bittner, M. *et al.* *Nature* **406**, 536–540 (2000).
3. Lockhart, D. J. & Winzler, E. A. *Nature* **405**, 827–836 (2000).
4. Berns, A. *Nature* **403**, 491–492 (2000).
5. Silletti, S., Paku, S. & Raz, A. *Int. J. Cancer* **76**, 120–128 (1998).
6. Chen, H., Bernstein, B. W. & Bamberg, J. R. *Trends Biochem. Sci.* **25**, 19–23 (2000).
7. Hall, A. *Science* **279**, 509–514 (1998).
8. Ridley, A. J. in *GTPases: Frontiers in Molecular Biology* (ed. Hall, A.) 89–136 (Oxford Univ. Press, 2000).
9. Itoh, K. *et al.* *Nature Med.* **5**, 221–225 (1999).
10. Allen, W. E., Zicha, D., Ridley, A. J. & Jones, G. E. *J. Cell Biol.* **141**, 1147–1157 (1998).
11. Stam, J. C. *et al.* *EMBO J.* **17**, 4066–4074 (1998).
12. Banyard, J. *et al.* *Oncogene* **19**, 580–591 (2000).
13. Lin, R., Cerione, R. A. & Manor, D. *J. Biol. Chem.* **274**, 23633–23641 (1999).

High-temperature superconductors

Some vortices like it hot

Patrick A. Lee

The discovery of high-temperature superconductivity, in a family of compounds containing copper–oxygen layers¹, triggered intense research into these copper oxides. It has become increasingly clear that the high-temperature superconductors are not just conventional superconductors with a higher superconducting-transition temperature, T_c — the point at which electrons pair up and begin to flow without resistance — but are fundamentally different beasts. On page 486 of this issue, Xu *et al.*² report that some aspects of superconductivity may reveal themselves at temperatures much higher than T_c . This work confirms how different the copper-oxide superconductors are, and how little we understand them at present.

Most significant among the differences is that, whereas conventional superconductors develop from the metallic state, the copper oxides start out life as insulators. Charge carriers are introduced into the insulator by the process of doping. For example, one starts with the insulator La_2CuO_4 and replaces some of the La^{3+} by Sr^{2+} , yielding $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. The result is that x electrons per copper atom are removed from the copper-oxide plane. When x is between 0.05 and 0.25, the material is superconducting, with the highest T_c around $x=0.15$. Moreover, La_2CuO_4 itself is an unusual insulator, called a Mott insulator. It is insulating by virtue of the strong repulsion between electrons occupying the same crystal lattice sites. So the problem with understanding high- T_c superconductivity is synonymous with the problem of understanding doping in a Mott insulator, a difficult issue in condensed-matter physics that has received a lot of attention only in the past decade.

For a superconductor that grows out of

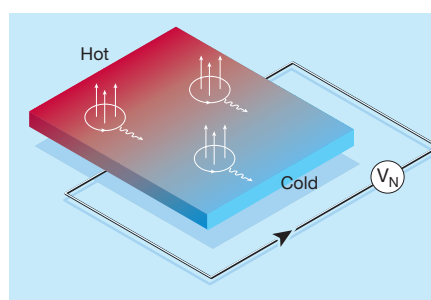


Figure 1 How to measure the Nernst effect due to vortices in superconductors. The vortices drift from hot to cold, carrying magnetic field lines with them, and generate a Nernst voltage, V_N . A large Nernst effect observed by Xu *et al.*² at temperatures above the transition temperature, T_c , in high-temperature superconductors suggests the presence of these vortices in the normal metallic state.

an insulator, one might expect the most unusual behaviour to be associated with low doping levels. This is indeed the case. Whereas the transition temperature, T_c , decreases (roughly linearly) with decreasing x , the ‘superconducting energy gap’, which characterizes the energy spectrum of the superconducting electrons, increases. This is in contrast to conventional superconductors, in which the ratio of T_c to energy gap is constant. Furthermore, the energy gap in copper oxides appears to persist to a temperature T^* , which is much higher than T_c , and also increases with decreasing x . This gap-like behaviour above T_c has been seen in photoemission experiments. In addition, when a material becomes superconducting, the electrons pair up into a so-called spin-singlet state and exhibit an activation gap in the spin excitation. Such an activation gap is

also seen above T_c in the copper oxides. The temperature range between T_c and T^* has been referred to as the pseudogap regime.

There are two schools of thought on the origin of the pseudogap. One is that the pseudogap state is fluctuating superconductivity — that is, over short distance and short timescales the system is a superconductor, but it cannot maintain such long-range behaviour because of fluctuations in the phase of the quantum wavefunction that describes the superconducting electron pairs. In particular, Emery and Kivelson³ pointed out that the parameter that controls the phase fluctuations is the superfluid density, ρ_s , which is proportional to x for a doped Mott insulator. For small x , phase fluctuations are strong and give rise to a relatively low T_c , which is proportional to x . However, above T_c the system remains superconducting at a short distance scale and exhibits many of the properties characteristic of a superconductor, such as the energy gap and the gap in the spin excitations. Another line of thought suggests that, because a copper oxide is close to being a Mott insulator, the electrons are mostly localized at the crystal lattice sites. The spins of these localized electrons form a singlet state, thus accounting for the pseudogap behaviour. In this view the pseudogap state is quite different from the superconducting state in that it mainly involves the spin of the electrons⁴.

Until now there has been no direct evidence for superconducting phase fluctuations in the pseudogap state. Phase fluctuations in a thin ‘two-dimensional’ superconductor are a well-researched subject, and are described by Berezinskii–Kosterlitz–Thouless (BKT) theory^{5,6}. The destruction of the superconducting long-range order is triggered by the appearance of free vortices. A vortex is an excitation of the superconducting current consisting of a core of normal metal around which the phase angle of the electron-pair wavefunction changes by 2π . Magnetic field lines can penetrate the superconductor through the vortices. By studying a.c. conductivity, researchers have shown that copper oxides with low doping appear to be described by the standard BKT theory⁷. Above T_c (in this particular experiment, 74 K), vortices multiply, so the fluctuation regime, although large compared with conventional superconductors, extends only up to temperatures of 100 K, well below the 250 K associated with the pseudogap.

In contrast, Xu *et al.*² study the Nernst effect, which is particularly sensitive to the existence of vortices. To observe the Nernst effect they measure the voltage transverse to a thermal gradient in the presence of a magnetic field perpendicular to the superconducting film (Fig. 1). The magnetic fields are tied to the vortices, which drift along the thermal gradient. A moving vortex then induces a transverse voltage. What Xu and

colleagues discover is that the Nernst signal persists well above T_c up to a temperature (T_{onset}) of 150 K. Furthermore, whereas T_c decreases with decreasing doping, T_{onset} increases, with the same trend as the T^* associated with the pseudogap. This is taken as evidence that vortex-like objects exist up to a temperature of 150 K, in samples where T_c ranges from 30 K to essentially zero.

The data are so striking and unexpected that the results need to be interpreted carefully. According to previous a.c. conductivity measurements⁷, vortices rapidly multiply above T_c . But eventually their cores start to overlap and the notion of vortices loses its meaning. This appears to happen about 30 K above T_c and seems at odds with the conclusion of Xu and colleagues. Admittedly, the two experiments look at different families of copper oxides, and the apparent discrepancy needs to be resolved by performing both experiments on similar samples. The Nernst measurement is possibly more sensitive to vortices than the a.c. method, but the measurement of T_{onset} relies on the assumption that the normal-state Nernst effect is

small and temperature independent. Although this is well established for conventional metals, it is worth recalling that all charge-transport properties are unusual in the underdoped copper oxides. So the meaning of T_{onset} must be viewed with some caution.

Nonetheless, these data clearly indicate strong vortex-like behaviour up to 100 K when the superconducting T_c is less than 30 K. Despite 100 K being much lower than the temperature scale associated with the pseudogap, the existence of such a large region of fluctuating superconductivity is a surprising finding that adds to the mystery of the high- T_c superconductors. ■

Patrick A. Lee is in the Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
e-mail: pal@mit.edu

1. Bednorz, J. G. & Müller, K. A. *Z. Phys. B* **64**, 189–193 (1986).
2. Xu, Z. A., Ong, N. P., Wang, Y., Kakeshita, T. & Uchida, S. *Nature* **406**, 486–488 (2000).
3. Emery, V. & Kivelson, S. *Nature* **374**, 434–437 (1995).
4. Wen, X. G. & Lee, P. A. *Phys. Rev. Lett.* **76**, 503–506 (1996).
5. Berezinskii, V. L. *Sov. Phys. JETP* **32**, 493–500 (1970).
6. Kosterlitz, J. M. & Thouless, D. J. *J. Phys. C* **6**, 1181–1203 (1973).
7. Corson, J. et al. *Nature* **393**, 221–223 (1999).

independent of variation in the rates of sediment accumulation, which is the weak point of many other productivity proxies.

Loubere concludes that the EEP has not responded uniformly to climate factors: one region (the Southern Equatorial Current area) exhibited lower productivity and another (the equatorial area) higher productivity during the last glacial. Because local winds would probably deposit iron uniformly throughout the whole EEP region, he suggests that iron fertilization is unlikely to be the primary control on productivity.

The productivity reconstruction implied by the foraminiferal analysis is straightforward and consistent with some (though not all) studies in this region^{8,9}. So Loubere has staged an impressive attack on the idea that iron has a global role in regulating marine productivity. But there are questions to be answered before accepting the death of the iron hypothesis.

Iron might indeed increase productivity in the sunlit surface waters, but that will not necessarily augment the amount of carbon removed to depth. Moreover, iron will increase primary productivity only when it is the ultimate limiting nutrient. If the supply of other nutrients (such as phosphate, nitrate or silica) is low, or other processes (such as water circulation and mixing patterns, or food-web dynamics) limit productivity, then extra iron will not help much. In the specific area of the Southern Equatorial Current, within the EEP, iron might not be limiting, so other processes or nutrients might be more important in regulating productivity.

On the other hand, iron could still be the key nutrient in other regions within the equatorial Pacific, including the south tropical Pacific⁶, the Panama Basin¹⁰ and the central¹¹ and western¹² equatorial Pacific, as well as in other oceans (the Atlantic and possibly the Southern Ocean). So the regional variability might indicate that, although iron was added to the EEP region during glacial times, the ultimate processes controlling productivity in each sub-zone were different.

A striking point brought out by Loubere's results is the evidence for small-scale variability in productivity, and thus nutrient distribution, in the EEP during the last glacial. This suggests that data obtained from a few sediment cores cannot be easily interpolated and used to interpret global or even regional environmental change — as Herguera¹² has pointed out, looking at a few tiles of a mosaic leaves us far from able to say what it looked like.

The underlying assumption with palaeo-oceanographic data obtained from only a few sites (cores) is that they represent regional conditions and can be extrapolated laterally. Yet high-resolution time series of regional oceanographic observations indicate that mesoscale variability in both space (tens of kilometres) and time (interannual and

Global change

Iron uncertainty

Adina Paytan

In these pages last week, there was an account¹ of a new explanation² for the lower levels of CO₂ in the atmosphere during the last glaciation (about 130,000 to 20,000 years ago) compared with the ensuing interglacial³. Given that CO₂ remains one of the main determinants of global warming, the matter is of obvious interest. The explanation held that it was the high availability of silica, rather than of iron⁴, that induced phytoplankton growth during the last glaciation. Results described by Loubere (page 497 of this issue⁵) show a different picture again, at least for the eastern equatorial Pacific Ocean (EEP); see the map on page 498. The EEP is an important area in terms of the high levels of both new ocean biological productivity (up to half of the global total) and ocean–atmosphere exchange of CO₂.

Here I will concentrate on the iron-fertilization hypothesis, which for several years has been at the centre of debate among those studying climate change. The idea is that during glacial periods the land was drier and the winds stronger, meaning that more iron-containing dust was blown over the oceans. Iron is a limiting nutrient for phytoplankton photosynthesis, so its greater availability increases productivity in the ocean surface⁴. Increased productivity results in greater incorporation of CO₂ into organic matter through photosynthesis. When the phytoplankton die or get

eaten, some of their remains sink to the ocean bottom, effectively removing CO₂ from contact with the atmosphere for a long time. The result, during glacial times, was lower concentrations of CO₂ in the atmosphere.

Over the past 50 years, oceanographers have used a whole range of indicators (proxies) to try to determine the difference in surface biological productivity between glacial and interglacial periods. Data reported so far for the EEP have produced equivocal results, and there is no consensus as to the direction or degree of any productivity change⁶. In an attempt to resolve the question, Loubere⁵ has analysed assemblages of bottom-dwelling organisms known as foraminifera in sediment cores from different regions of the EEP. He then used these assemblages to infer surface productivity at different times in the past.

For the most part, organisms dwelling at the bottom of the sea depend entirely on the rain of organic matter — food particles — falling from above. So it is likely that the abundance and distribution of such organisms will be controlled by the amount of organic material they receive. Other variables such as substrate type, oxygen availability and currents will also have an influence. But in the ocean today, bottom-living foraminifera do indeed provide a highly sensitive record of environmental changes associated with the amount of food available⁷. This productivity proxy is

decadal) is considerable (J. A. McGowan, personal communication). Reliable, high-resolution historical maps of productivity, similar to those derived from present-day satellite imaging data, would help a great deal in understanding the distribution of, and controls on, past productivity. Unfortunately such maps are unlikely to be available for sometime.

The work by Loubere⁵ adds important information about palaeoproductivity in the EEP. But the apparent conflict in glacial productivity reconstruction using different proxies remains, and it will have to be settled. Because several processes affect each of these proxies, the signals obtained may reflect a complex interaction between those processes, rather than productivity alone. At any given place, a multi-proxy approach is required, as is a thorough examination of all the assumptions underlying, and processes affecting, each of the proxies used. Moreover, independent proxies for iron availability or lack of it

— such as the composition of specific organic compounds, the ratios of silicate to nitrate and phosphate, or some iron-concentration proxy preserved in mineral phases — are needed to get a true understanding of iron's role in regulating productivity. ■

Adina Paytan is in the Department of Geological and Environmental Sciences, Braun Hall (320) 207, Stanford University, Stanford, California 94305-2115, USA.

e-mail: apaytan@pangea.stanford.edu

1. Tréguer, P. & Pondaven, P. *Nature* **406**, 358–359 (2000).
2. Harrison, K. G. *Paleoceanography* **15**, 292–298 (2000).
3. Delmas, R. J., Ascencio, J.-M. & Legrand, M. *Nature* **284**, 155–157 (1980).
4. Martin, J. H. *Paleoceanography* **5**, 1–13 (1990).
5. Loubere, P. *Nature* **406**, 497–500 (2000).
6. Loubere, P. *Mar. Micropaleontol.* **37**, 173–198 (1999).
7. Loubere, P. & Fariduddin, M. *Glob. Biogeochem. Cycles* **13**, 115–134 (1999).
8. Lyle, M. *et al. Paleoceanography* **3**, 39–59 (1988).
9. Herguera, J. *Mar. Micropaleontol.* (in the press).
10. Yang, Y., Elderfield, H. & Pedersen, T. *Geology* **23**, 115–118 (1995).
11. Prahil, F. *et al. Paleoceanography* **4**, 495–510 (1989).
12. Herguera, J. & Berger, W. *Geology* **19**, 1173–1176 (1991).

Bacterial genomics

Treasure trove for cholera research

Matthew K. Waldor and Debabrata RayChaudhuri

Cholera has afflicted human beings and shaped human history for over two millennia. The causative infectious agent of this diarrhoeal disease, the comma-shaped bacterium *Vibrio cholerae* (Fig. 1), colonizes the mucosal surface of the human small intestine and secretes cholera toxin. The toxin stimulates secretion of water and electrolytes by the cells of the small intestine, leading to the severe watery diarrhoea that is characteristic of cholera¹. Unlike most bacterial agents of diarrhoea, *V. cholerae* tends to survive in aquatic environments during periods between epidemics, can give rise to dramatic outbreaks of disease, and has spread from its endemic focus in Asia, causing seven global 'pandemics' so far.

The study of this disease has a remarkable history, having spawned central tenets in the fields of epidemiology (starting with John Snow's famous tracing of an outbreak to a water pump in Broad Street, London, in 1854), treatment for diarrhoeal disease, microbial ecology and pathogenesis, and mucosal immunology. The complete genome sequence of a representative isolate from the ongoing seventh pandemic of cholera is described by Heidelberg *et al.*² on page 477 of this issue, and promises to usher in an exciting era of post-genomic investigation of this organism.

The *V. cholerae* genome consists of two circular chromosomes (Fig. 2), the larger comprising some 2.96 million base pairs (megabases) and the smaller being of roughly

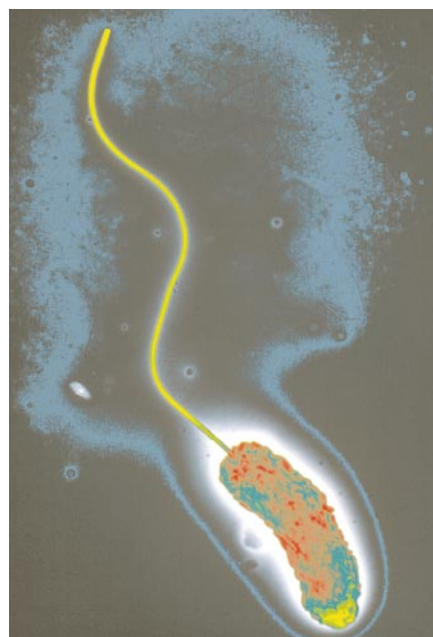


Figure 1 *Vibrio cholerae*, the bacterium that causes cholera. The epidemiology and biology of cholera infections have been studied for centuries. Now, with the publication of the bacterium's genome sequence², comes the opportunity for post-genomic analyses. Magnification approximately $\times 10,000$.

1.07 megabases². Both chromosomes have roughly the same amount of guanine plus cytosine nucleotides, constituting about 47% of the genome. This suggests to us that these two units of replication (replicons) have exist-

ed together for a long period, even though Heidelberg *et al.*² suggest disparate evolutionary origins for these chromosomes.

The larger chromosome, chromosome I, contains a preponderance of genes dedicated to essential cell functions such as DNA replication, cell division, gene transcription, protein translation and cell-wall biosynthesis. It also has most of the genetic loci that are associated with virulence (Fig. 2), many of which have been acquired from other species. For example, a major determinant of pathogenicity, cholera toxin³, is encoded in the genome of a virus that integrates into the *V. cholerae* genome⁴ on chromosome I. A factor essential for colonization of the intestine, toxin-coregulated pilus⁵, is encoded on chromosome I in a 'Vibrio pathogenicity island' found in *V. cholerae* strains that cause pandemics⁶. The expression of genes involved in virulence is regulated by proteins such as ToxR and ToxT^{7,8}, both of which, along with several other such regulators, are also encoded on the larger chromosome (Fig. 2).

A longstanding view, based on studies of *Escherichia coli*, held that bacterial cells each contain a single circle of chromosomal DNA. But this view has changed with the discovery of an increasing number of microorganisms that contain more than one chromosome, in circular and linear configurations^{9,10}. The presence of essential genes is thought to be a defining feature of chromosomes⁹. Several essential genes, including those encoding the ribosomal proteins L20 and L35, are present only on chromosome II, confirming its status as a bona fide chromosome. The presence of essential genes on chromosome II also indicates that it is probably inherently stable — it is unlikely that essential functions would be delegated to an unstable replicon during the course of evolution. Genes from all 16 categories used to subdivide cellular functions² occur on chromosome II, although, for eight of these classes, it contains significantly fewer genes than chromosome I. Represented prominently on chromosome II are genes involved in the transport of sugars, metal ions and anions, in the metabolism of sugars and energy, in two-component signal transduction, and in DNA repair.

Why and how does genetic information become distributed onto separate replicons? Heidelberg *et al.*² argue that chromosome II of *V. cholerae* was acquired as a plasmid, a circular piece of DNA that replicates autonomously. But we cannot rule out the possibility that the small chromosome arose by excision from a single large ancestral genome. One would imagine that, by distributing the genome content among more than one chromosome, a bacterium could achieve faster genome duplication and cell division, with the possible upshot being a growth advantage and an increased ability to survive in a competitive environment. Heidelberg *et al.* found 105 genes with copies on both

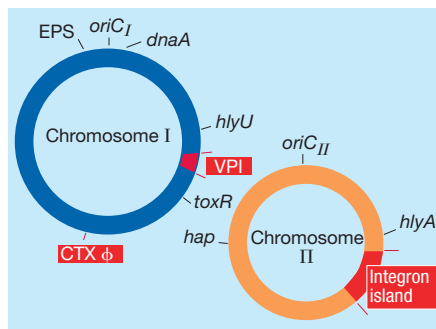


Figure 2 Distribution on the two *Vibrio cholerae* chromosomes of key genetic loci required for virulence². The DNA sequences acquired from other species, such as the *Vibrio* pathogenicity island (VPI)⁶, the temperate bacteriophage CTXφ⁴ and the integrin island², are shown in red. The VPI encodes the essential intestinal colonization factor toxin-coregulated pilus⁵ and a virulence regulator, ToxT^{7,8}. CTXφ encodes cholera toxin⁴. The integrin island is a gene-capture system yet to be associated with virulence. The pathogenicity determinants that have not been recently acquired (black) include: *toxR*, which encodes a transcriptional regulator of several virulence loci; *hlyU*, which encodes a transcriptional regulator that controls the expression of the haemolysin-encoding gene *hlyA*; and EPS, a protein-secretion system that is required for the export of CTXφ, cholera toxin and several proteases, including Hap. The dissimilar origins of replication on chromosomes I and II are designated as *oriC_I* and *oriC_{II}*, respectively. Also shown is *dnaA*, which encodes a protein that is likely to initiate replication of chromosome I.

V. cholerae chromosomes, reflecting the occurrence of considerable interchromosomal recombination events. But the persistence of the two independent replicons underscores the strong selective pressure for maintaining two stable chromosomes.

The presence of more than one chromosome is a universal feature of eukaryotic cells. The *V. cholerae* genome sequence now offers bacteriologists an opportunity, in a genetically tractable organism, to study the crisscrossing functional and regulatory interactions that occur between genes and their products from the two chromosomes. Gene products from the two chromosomes interact functionally in metabolic pathways, nutrient transport systems, and biosynthesis of the flagellum, the whip-like structure that acts as a propeller to move the organism about. An example of interchromosomal regulatory interactions is offered by RpoS, a transcription factor specific to the stationary phase of the bacterium's life cycle and encoded on chromosome I. RpoS regulates the expression of several genes on both chromosomes.

Moreover, there must be specific — perhaps novel — mechanisms to coordinate replication and segregation of the two chromosomes into progeny *V. cholerae* cells. The point at which DNA replication starts

on chromosome I (*oriC_I*) is similar to the canonical replicative origin (*oriC*) from *E. coli*. The putative origin on the smaller chromosome is different, and was assigned by Heidelberg *et al.* solely on the basis of nucleotide-composition analysis. Interestingly, this putative origin of replication, unlike that on chromosome I, lacks recognition sites for the replication-initiator protein DnaA and for Dam methyltransferase, which regulates the initiation of replication by methylating adenine nucleotide bases in *oriC*. It seems that the replication of the two chromosomes may well be controlled by different mechanisms.

All in all, the *V. cholerae* genome sequence will be a treasure trove for researchers who wish to study the replication and segregation of the two chromosomes and to analyse the functions¹¹ of the many putative genetic loci on the chromosomes. It will also inspire a comparative exploration of the genomic alphabets of different *V. cholerae* subtypes and *Vibrio* species, with the aim being to understand the evolution and the bases of pathogenicity and environmental persis-

tence of these microorganisms. Last but not least, the genome sequence should aid in the development of new vaccines and medicines to tackle the ancient scourge of cholera. ■

Matthew K. Waldor is at the Howard Hughes Medical Institute and the Department of Medicine, Tufts New England Medical Center, NEMC #041, 750 Washington Street, Boston, Massachusetts 02111, USA.

e-mail: mwaldor@lifespan.org

Debabrata RayChaudhuri is in the Department of Molecular Biology and Microbiology, Tufts University School of Medicine, 136 Harrison Avenue, Boston, Massachusetts 02111, USA.

e-mail: draychau@opal.tufts.edu

1. De, S. N. *Nature* **183**, 1533–1534 (1959).
2. Heidelberg, J. F. *et al.* *Nature* **406**, 477–483 (2000).
3. Mekalanos, J. J. *et al.* *Nature* **306**, 551–557 (1983).
4. Waldor, M. K. & Mekalanos, J. J. *Science* **272**, 1910–1914 (1996).
5. Taylor, R. K. *et al.* *Proc. Natl Acad. Sci. USA* **84**, 2833–2837 (1987).
6. Karalis, D. K. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 3134–3139 (1998).
7. Skorupski, K. & Taylor, R. K. *Mol. Microbiol.* **25**, 1003–1009 (1997).
8. Lee, S. H. *et al.* *Cell* **99**, 625–634 (1999).
9. Kolsto, A.-B. *Mol. Microbiol.* **24**, 241–248 (1997).
10. Yamaichi, Y. *et al.* *Mol. Microbiol.* **31**, 1513–1521 (1999).
11. Judson, N. & Mekalanos, J. J. *Nature Biotechnol.* **18**, 740–745 (2000).

Geophysics

Taking the Earth's temperature

Ian Jackson

The slow overturning of the Earth's interior, which we see on the surface in the motions of the tectonic plates, is driven by differences in temperature and composition of the interior. Direct measurements of temperatures in mines and boreholes (to a maximum depth of about 10 km) sample only the upper crust. Some information about temperatures at greater depths (to 200 km) in the upper mantle is provided by the analysis of minerals from rock fragments carried to the surface by volcanic magmas. But much of our knowledge of the structure of the Earth's interior and, indirectly, its composition and temperature, comes from seismology — the study of seismic waves produced by earthquakes. Much of the variation in seismic wave speeds in the mantle is probably due to variations in temperature, but attempts to model these variations have been few and far between. Now, however, better seismic models and experimental data on the physical properties of mantle rocks make such analysis both feasible and rewarding. The state of the art in the interpretation of seismic tomography is described in a paper by Goes *et al.*¹ in the *Journal of Geophysical Research*.

Seismic tomography is loosely analogous with the more familiar 'CAT scan' of medical testing. In the CAT scan, the subject is uniformly illuminated with X-rays that travel along simple straight-line paths to an array

of closely spaced detectors. An image of the internal structure responsible for greater or lesser absorption along particular paths is created from the detector outputs. Unlike X-rays, seismic rays are strongly bent (refracted) by variations in wave speed — so they travel from an earthquake source back to the Earth's surface along curved paths. Despite this added complication and the non-uniform distribution of earthquake sources and receivers, geophysicists have obtained reliable three-dimensional images of the Earth's interior on regional and global scales.

There are two types of seismic wave that can pass through the Earth: primary and secondary waves. Primary (P or compressional) waves alternately compress and dilate the matter they travel through (either rock or liquid), and are analogous to sound waves in a fluid medium. Secondary (S or shear) waves travel more slowly than primary waves through rock, but cannot travel through liquid at all. Both P and S waves travel more slowly through hotter material and are also refracted by gradients in chemical composition. The speeds of P and S waves have been measured in minerals and rocks of known chemical composition under controlled conditions of pressure and temperature in the laboratory.

Most experimental data on seismic wave speeds have been obtained at the very high

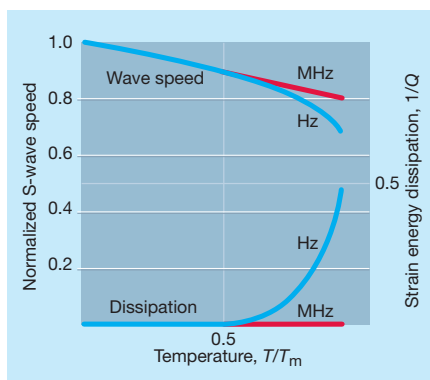


Figure 1 The effects of temperature on S-wave speed and dissipation of strain energy (after ref. 13). As a consequence of thermally activated viscoelastic relaxation processes involving the motion of crystal defects, the properties measured at ultrasonic frequencies (MHz) and at seismic frequencies (Hz) diverge as the temperature rises towards the melting temperature, T_m . Q is quality factor.

frequencies (MHz to GHz) associated with ultrasonic and optoacoustic techniques, and have provided vital information about the pressure- and temperature-dependence of wave speeds. But applying such information to typical seismic frequencies in the Earth (in the range 0.01–1 Hz) is complicated by the effects of viscoelasticity, especially on S-wave speed. At sufficiently high temperatures and low frequencies, the response of rocks to shear stress involves not only the usual elastic response, but also both recoverable and irrecoverable viscoelastic processes. Such viscoelasticity leads to dissipation of strain energy and frequency dependence (dispersion) of the wave speed. Experimental studies of the viscoelastic behaviour of rocks at seismic frequencies^{2,3} show that S-wave speeds are indeed more sensitive to temperature changes than they are at higher frequencies, as predicted by the theory of linear viscoelasticity^{4,5} (Fig. 1).

In their analysis of the available data on mantle rocks, Goes *et al.*¹ conclude, in agreement with previous studies^{6,7}, that the upper-mantle seismic wave speeds are much more sensitive to changes in temperature than changes in rock composition. Accordingly, they have analysed models of the P-wave⁸ and S-wave⁹ speeds for the European upper mantle to obtain a picture of the temperature variation beneath Europe. This analysis uses a strategy broadly similar to those of previous studies^{6,10}, but differs from them by including the effects of viscoelasticity in separate interpretations of P- and S-wave models.

Goes and colleagues find that the amplitude of the variation in S-wave speed is consistently greater than that for P-wave speed, suggesting that the temperature sensitivity of the S waves may be enhanced by viscoelastic relaxation. This view is confirmed by the impressive consistency between tem-

perature models inferred separately from the P- and S-wave data, at all but the shallowest depth (55 km), where the analysis may be compromised by the influence of crustal structure and variations in composition.

Moreover, the mantle temperatures calculated by Goes *et al.* agree reasonably well with independent estimates from heat flow and general geological considerations. At 100-km depth, temperatures beneath the Russian platform are about 300 °C lower than average temperatures for the bulk of Western Europe. Here, temperatures can be ~100 °C higher below areas of recent volcanism (such as the Rhenish Massif and Massif Central) or areas of active stretching of the crust and uppermost mantle. Beneath such extensional basins, the higher temperatures found using S-wave rather than P-wave data (especially at depths of less than 100 km) suggest the possibility of localized melting.

Goes and colleagues' analysis convincingly demonstrates that the variability of seismic wave speeds in the European upper mantle is primarily of thermal origin. A simple but robust modelling of the influence of viscoelasticity accounts well for the greater variation in S- than in P-wave speed — resulting in broadly consistent temperatures from S- and P-wave models. Further progress can be expected as tomographic models are refined. For example, an impressive correlation between P-wave speed and energy dissipation in the upper mantle beneath the Tonga–Fiji region, plausibly of mainly thermal origin, has recently been found¹¹. Moreover, laboratory data are yielding ever tighter constraints on viscoelastic behaviour in the solid state and on the influence of partial melting^{3,12}. These advances will mean better interpretations of three-dimensional seismic structure and provide further insight into the temperature and composition of the upper mantle. ■

Ian Jackson is at the Research School of Earth Sciences, Australian National University, Canberra, ACT 0200, Australia.
e-mail: ian.jackson@anu.edu.au

- Goes, S., Govers, R. & Vacher, P. J. *Geophys. Res.* **105**, 11153–11169 (2000).
- Berckhemer, H., Kampfmann, W., Aulbach, E. & Schmeling, H. *Phys. Earth Planet. Inter.* **29**, 30–41 (1982).
- Jackson, I. *Geophys. Monogr.* **117**, 265–289 (2000).
- Minster, J. B. & Anderson, D. L. *Phil. Trans. R. Soc. Lond.* **299**, 319–356 (1981).
- Karato, S. *Geophys. Res. Lett.* **20**, 1623–1626 (1993).
- Jordan, T. H. in *The Mantle Sample: Inclusions in Kimberlites and Other Volcanics* (eds Boyd, F. R. & Meyer, H. O. A.) 1–14 (AGU, Washington, 1979).
- Sobolev, S. V. *et al. Earth Planet. Sci. Lett.* **139**, 147–163 (1996).
- Bijwaard, H., Spakman, W. & Engdahl, E. R. *J. Geophys. Res.* **103**, 30055–30078 (1998).
- Marquering, H. & Snieder, R. *Geophys. J. Int.* **124**, 283–304 (1996).
- Furlong, K. P., Spakman, W. & Wortel, M. J. *R. Tectonophysics* **224**, 107–117 (1995).
- Roth, E. G., Wiens, D. A. & Zhao, D. *Geophys. Res. Lett.* **27**, 601–604 (2000).
- Gribb, T. T. & Cooper, R. F. *J. Geophys. Res.* **103**, 27267–27279 (1998).
- Goetze, C. in *High-Pressure Research: Applications in Geophysics* (eds Manghni, M. H. & Akimoto, S.) 3–23 (Academic, New York, 1977).

Daedalus

The soda solution

Sodium carbonate, ordinary washing soda, melts to a most powerful reagent. It dissolves all minerals, even silica — indeed, glass is made by dissolving silica and other minerals in molten sodium carbonate. Furthermore, it attacks organic matter ferociously. One process for disposing of halocarbon plastics, old tyres, leather scraps, and so on, submerges them in a sodium carbonate melt and blows in a trickle of air. The organics react to a useful fuel gas containing carbon monoxide and hydrogen. Their halogens and sulphur stay in the melt as halides and sulphide. Sadly, the contaminated carbonate is troublesome to recycle. It has to be dissolved in water, chemically purified, then re-evaporated and remelted.

To simplify this elegant process, Daedalus is extending it to fuel minerals as well. He points out the vast amount of fuel locked up in awkward and uncooperative geological forms — tar sands, low-grade shales, alluvial carbides, and so on. So DREADCO chemical engineers are now equipping a molten-carbonate rubbish-reducer with a second, fuel-processing stage. After digesting its charge of organic rubbish, the melt flows into another under-aerated vessel, where tar sand or shale is added. Its organic content burns incompletely to fuel gas, as before; its mineral content, silica or silicates, dissolves in the carbonate. The silica displaces the halides and sulphide, and the carbonate of the soda, as gases which can be scrubbed out of the fuel-gas downstream. The transformed melt consists of mixed silicates of sodium, calcium and the other metals in the fuel mineral. In fact it is simply molten glass, a useful product in its own right.

To get rid of all our rubbish, and make a useful contribution to our fuel-gas supply, the new process would have to be worked on a very large scale. Luckily, sodium carbonate itself is a natural mineral, available in quantity; but the resulting vast output of rather poor and probably discoloured glass would saturate the existing market. But Daedalus recalls a previous invention of his, a foamed brick. This useful building component combined lightness with heat insulation. A foamed glass brick would add translucency to its virtues. So Daedalus will aerate his molten glass and cast it into bricks. They should sweep the building trade. A process which turns rubbish and intractable minerals into fuel gas and housing should usefully boost DREADCO's green credentials.

David Jones

Obituary

Seymour S. Kety (1915–2000)

Seymour Kety, who died peacefully in his sleep on 24 May at the age of 84, was at the centre of advances in understanding mental disorders for much of the twentieth century. He was a key figure in the development of biological research in psychiatry; he established a tradition of excellence as the scientific director of the US National Institute of Mental Health (NIMH); and he trained several generations of leading mental health investigators.

In 1945, Kety developed the first quantitative and reproducible method for measuring cerebral blood flow. The implications of that simple statement will escape those not old enough to recall the bizarrely unreliable techniques previously used. Before the new method was introduced, oxygen tension was measured twice in blood sampled from the internal carotid artery and the internal jugular vein; if blood flow was assumed to be constant during that interval, changes in the differential between arterial and venous oxygen levels reflected metabolism; if cerebral metabolism was taken as unchanged, the difference could only reflect changes in blood flow. But which of the two was the case? There was no way to know. Kety used nitrous oxide, an inert but soluble gas, to measure blood flow (by applying the Fick principle). Once blood flow had been calculated, changes in arteriovenous oxygen tension provided an index of metabolism.

Kety and his colleagues applied the method to healthy volunteers, and to patients with various ailments — essential hypertension, diabetic acidosis and coma, increased intracranial pressure and senile dementia. Major differences were evident in these patients. But blood flow and oxygen consumption in people with schizophrenia were entirely within the normal range. Kety knew that such findings did not rule out the possibility of biologically important changes in particular brain regions of schizophrenics. Thus, measuring differences in regional blood flow became his next objective.

Building on Bohr's derivation for the exchange of oxygen at the capillary in a steady state, he developed an expression for the equilibration of an inert tracer in terms of capillary geometry, diffusion coefficients and perfusion, and used it to construct an equation for calculating regional cerebral flow. By using radioactive tracers, tissue concentrations could be measured by autoradiography in animals and by



Key figure in the development of biological research in psychiatry

external counters in man. At NIMH, he was joined by William Landau, Louis Sokoloff, Lewis Roland and Walter Freygang; by 1955, they were able to report blood flow in 28 brain structures in animals.

Kety established an uncommonly productive programme at NIMH to investigate brain biochemistry and physiology. Under his guidance, Julius Axelrod carried out his Nobel prize-winning work on catecholamine metabolism; Louis Sokoloff developed his spectacularly successful deoxyglucose method for measuring local cerebral metabolism; and Marcus Raichle applied these techniques to the study of cognitive function by positron emission tomography.

Equally seminal contributions arose from Kety's studies into the roots of schizophrenia. He used a national sample of adopted children who became schizophrenic to parse out genetic and environmental variables by comparing their biological and adoptive families. With David Rosenthal and Paul Wender at NIMH, and Fini Schulsinger in Copenhagen, he used the Danish Case Register as his sample source. Because mental illness predominated in the biological families, the study provided firm evidence for a genetic contribution to the transmission of schizophrenia. Few today doubt that the schizophrenias have a biological basis; at that time, the assertion was almost heresy.

Although Kety was a biologist, his comprehension of the genesis of mental disorders was far broader. In an era when psychiatry was being polarized between 'brainlessness' and 'mindlessness', he wrote a remarkable paper entitled "A biologist examines mind and behavior" (*Science* 132, 1861–1870; 1960). To highlight the need for multiple levels of inquiry, he invoked a powerful metaphor.

Imagine, he asked the reader, a civilization of high intelligence whose inhabitants had never seen a book. On discovering a library, they establish a scientific institute for studying books, one employing anatomists, physical chemists, molecular biologists, behavioural scientists and psychoanalysts. Each discipline discovers important facts (the structure of cellulose, the percentile frequency of collections of letters of varying length and so on). But the meaning of a 'book' continues to escape them. As he put it: "We do not always get closer to the truth as we slice and homogenize and isolate... a truer picture of the nervous system and behavior will emerge only from its study by a variety of disciplines and techniques, each with its own virtues and its own particular limitations."

My first contact with Kety was in 1944. He had just begun his academic career as an instructor in pharmacology at the Medical School of the University of Pennsylvania. I was a second-year medical student. His lectures were a model of clarity. Not only did they shed light on the obscure, but they made us aware of the frontiers of knowledge, the limits of understanding, and the importance of the unanswered and often unasked questions about the pathophysiology of disease.

In the year before his death, Kety received an Award for Special Achievement in Medical Science from the Lasker Foundation for a lifetime of contribution to neuroscience and for his "visionary leadership in mental health that ushered psychiatry into the molecular era". The citation says it all.

Those of us privileged to know Seymour Kety had our lives enriched by his mentorship, by his commitment to excellence and by his personal warmth. His surviving family include his wife, Dr Josephine Gross Kety, to whom he had been married for 60 years; a son, Lawrence; a daughter Roberta; and two grandchildren.

Leon Eisenberg

Leon Eisenberg is in the Department of Social Medicine, Harvard Medical School, 641 Huntington Avenue, 2nd Floor, Boston, Massachusetts 02115-6019, USA.
e-mail: leon_eisenberg@hms.harvard.edu

Oncogene inactivation in a mouse model

Tissue invasion by leukaemic cells is stalled by loading them with a designer ribozyme.

Chronic myelogenous leukaemia (CML) is a haematopoietic malignant disease associated with the expression of a chimaeric *BCR-ABL* gene^{1,2}. We have designed an allosterically controllable ribozyme that specifically cleaves *BCR-ABL* messenger RNA and induces apoptosis in cultured CML cells³, and here we test it as a possible treatment of CML in a mouse model. We find that this ribozyme completely inhibits tumour-cell infiltration in these mice. To our knowledge, this is the first application of an artificial, allosterically controllable enzyme in animals, opening up the possibility of using ribozyme technology in the treatment of CML.

The chimaeric *BCR-ABL* gene generates two different *BCR-ABL* mRNAs, known as K28 b3a2 and L6 b2a2, both of which yield p210^{BCR-ABL} on translation, an oncoprotein^{4,5} with tyrosine kinase activity² that causes malignant transformation of white blood cells in CML.

Although ribozyme-based gene therapy is arousing great interest, its application to disrupt chimaeric mRNA has been restricted by the lack of a specific cleavable site at the junction of the chimaeric gene. To disrupt *BCR-ABL* mRNA, a ribozyme must exclusively target the junction sequence, otherwise normal *ABL* mRNA that shares part of the chimaeric sequence will also be cleaved by the ribozyme, with resultant damage to host cells³.

We have previously generated a novel ribozyme ('maxizyme') that is dimeric^{3,6-8} and specifically cleaves *BCR-ABL* mRNA, inducing apoptosis in CML cells³. The maxizyme has sensor arms that recognize two target sequences that induce it to form a cavity for capturing Mg²⁺ ions essential for catalysis. The activity of the heterodimeric ribozyme can be allosterically controlled through one of the two substrate-binding regions³. The maxizyme remains in an inactive conformation in the presence of normal *ABL* mRNA or in the absence of the junction sequence³.

To test this technology *in vivo*, we embedded genes encoding the maxizyme downstream of genes for the human transfer RNA tRNA^{Val} (refs 3,7,8). We used a retroviral system for the expression of the maxizyme in leukaemic cells⁹, which involved subcloning two tRNA^{Val}-driven expression cassettes, corresponding to each component of the heterodimer, in tandem into the retroviral vector. A line of CML cells (BV173) was transduced either with a control vector in which the maxizyme sequence had been deleted, or with the

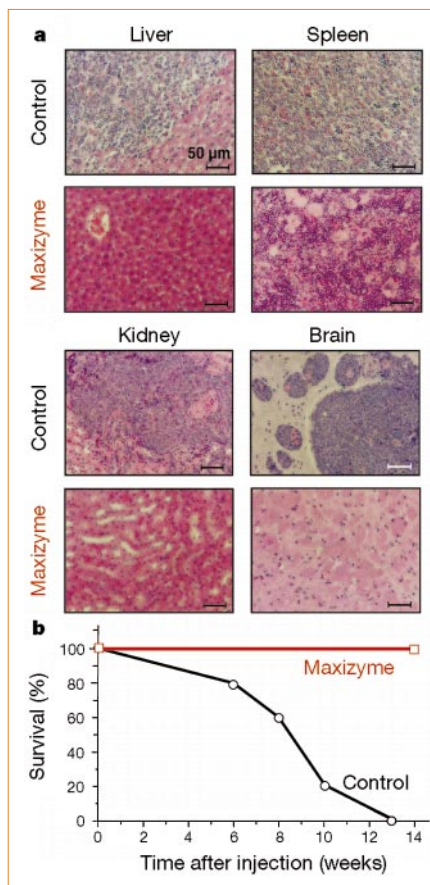


Figure 1 The antitumour effects of the maxizyme in a murine model of chronic myelogenous leukaemia (CML). **a**, Infiltration by BV173 leukaemic cells, as revealed by haematoxylin-and-eosin staining of livers, spleens, kidneys and brains of NOD-SCID mice injected with BV173 cells that had been transduced either with a control vector or with the maxizyme-expressing vector. In most cells stained with haematoxylin and eosin, the nucleus is blue and the cytoplasm pink, but the extent of staining is different for each tissue. Because infiltrated leukaemic cells almost replace normal cells in control spleen and disrupt its structure, the degree of staining is different from that of maxizyme-containing spleen. **b**, Reduction of tumorigenicity of BV173 cells *in vivo*. A total of 2×10^6 BV173 cells transduced with either control vector or maxizyme were selected by incubation with puromycin and injected into the tail vein of NOD-SCID mice (two independent experiments, with 4–8 animals per group in each experiment). The survival of animals was monitored daily for more than 20 weeks after inoculation: all control mice died within 13 weeks, whereas maxizyme-treated mice remained disease-free for the entire period of the investigation. Scale bars, 50 μ m.

maxizyme-encoding vector. We then injected 2×10^6 transduced BV173 cells into the tail veins of individual NOD-SCID mice. (These mice, produced by crossing SCID and non-obesity diabetic mice, are a powerful tool for growing human haematopoietic cells *in vivo*.) Survival of animals was monitored daily for 100 days after inocula-

tion. Eight weeks after inoculation, eight animals from each group were killed by cervical dislocation and their tissues examined.

The pathological changes found at necropsy were remarkably consistent in mice that had received control-transduced BV173 cells (Fig. 1a). The spleen was red, firm and enlarged, weighing two to three times more than the spleen from mice injected with maxizyme-transduced BV173 cells. Histological examination revealed leukaemic nodules, and the liver and kidney were enlarged and firm, with evidence of extensive metastasis of BV173 cells, which was particularly marked in the brain. By contrast, the spleens, livers, kidneys and brains of mice treated with maxizyme-transduced BV173 cells appeared normal (Fig. 1a).

The differences between the two groups of mice were reflected in their mortality rates (Fig. 1b). All of the mice injected with control BV173 cells died of diffuse leukaemia, confirmed at necropsy, 6–13 weeks afterwards (median survival time, 9 weeks), whereas mice injected with maxizyme-transduced BV173 cells remained healthy (Fig. 1b).

Our results indicate that each subunit of the maxizyme introduced by the retroviral vector is produced at the appropriate concentration to support dimerization *in vivo*. Also, the maxizyme apparently functions successfully in animals, cleaving *BCR-ABL* mRNA with exceptional efficiency.

At present, allogeneic transplantation is the only effective therapy for this type of leukaemia^{10,11}, with only half of all patients on average being eligible for this treatment because of limited donor availability and age restrictions. Our results raise the possibility that this maxizyme could be useful for purging bone marrow in cases of CML treated by autologous transplantation, when it would presumably reduce the incidence of relapse by decreasing the tumorigenicity of contaminating CML cells in the transplant¹².

We have also constructed five other maxizymes that successfully target other chimaeric genes¹³, suggesting that maxizymes could be a new class of powerful gene-inactivating agents that can cleave any type of chimaeric mRNA.

Tsuyoshi Tanabe†‡, Tomoko Kuwabara†‡, Masaki Warashina†‡, Kenzaburo Tani*, Kazunari Taira†§, Shigetaka Asano*

*Department of Hematology/Oncology, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8539, Japan

†National Institute for Advanced Interdisciplinary Research, 1-1-4 Higashi, Tsukuba Science City 305-8562, Japan

‡Institute of Applied Biochemistry, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba Science City 305-8572, Japan

§Department of Chemistry and Biotechnology, Graduate School of Engineering, University of Tokyo, Hongo, Tokyo 113-8656, Japan
e-mail: taira@chembio.t.u-tokyo.ac.jp

1. Bartram, C. R. *et al. Nature* **306**, 277–280 (1983).
2. Konopka, J. B., Watanabe, S. M. & Witte, O. N. *Cell* **37**, 1035–1042 (1984).
3. Kuwabara, T. *et al. Mol. Cell* **2**, 617–627 (1998).

Entomology

Home-site fidelity in migratory honeybees

Home-site fidelity is well known in migratory animals, but not in social insects. Here we show that colonies of the giant honeybee *Apis dorsata* are able to find the same nest location even after seasonal migration. As worker bees do not have first-hand knowledge of the old nest site, the swarms must be guided by some form of genetic mechanism.

Migratory vertebrates can often precisely locate their home sites after displacement or seasonal migrations. *A. dorsata* is migratory, following seasonal changes in flowering patterns^{1,2}. The existence of specific ‘bee trees’ that host colonies on a regular seasonal basis has led to the suggestion³ that these migrating colonies return to the same nesting location every year. We tested this idea by using DNA microsatellite genotyping of colonies sampled on the same bee trees over a three-year period.

Workers of five colonies (*n* = 12 each) at two bee trees (A and B, 2 km apart in Tenom, Malaysia) from five different branches (A1, A2, B1–3) were genotyped in 1995 (ref. 4) and 1997 using three DNA microsatellites^{5,6}. In 1996, only one colony occupied the nest sites and no bees were sampled. Queen and drone genotypes were derived from worker offspring⁶ and compared between the two years to test the identity of colonies at the same locations (Table 1).

Although only five colonies were sampled each year, population allele frequencies could be determined from 144 gametes because queens mate with many different males^{4,7}. All probabilities for random genotype matching⁸ computed from these allele frequencies were *P* < 0.01. The two samples from site B2 showed identical queen and drone genotypes, indicating that the same colony, headed by the same queen, had revisited the site. Two other pairs of colonies (A2 and B3) showed queen genotypes with one common allele at each locus, as expected for mother–daughter relationships. The other colonies sampled at the

4. Daley, G. Q., Van Etten, R. A. & Baltimore, D. *Science* **247**, 824–830 (1990).
5. Gishizky, M. L. & Witte, O. N. *Science* **256**, 836–839 (1992).
6. Amontov, S. & Taira, K. *J. Am. Chem. Soc.* **118**, 1624–1628 (1996).
7. Kuwabara, T. *et al. Nature Biotechnol.* **16**, 961–965 (1998).
8. Kuwabara, T., Warashina, M., Nakayama, A., Ohkawa, J. & Taira, K. *Proc. Natl Acad. Sci. USA* **96**, 1886–1891 (1999).
9. Onishi, M. *Exp. Hematol.* **24**, 324–329 (1996).
10. Snyder, D. S. & McGlave, P. B. *Hematol. Oncol. Clin. North Am.* **4**, 535–557 (1990).
11. McGlave, P. B. *et al. Blood* **81**, 543–550 (1993).
12. Deisseroth, A. B. *et al. Blood* **83**, 3068 (1994).
13. Tanabe, T. *et al. Biomacromolecules* **1**, 108–117 (2000).

same locations (A1, B1) showed no matching genotypes, as reported previously⁹.

These results indicate that one *A. dorsata* colony returned, with the same queen, to its home site after seasonal migration³. This also

shows that the lifespan of queens can exceed three years. In two other cases (A2, B3), putative daughter colonies were found at maternal sites after two years. How can colonies find exactly the same location when all workers have died¹⁰ and scouts do not have first-hand knowledge of the old nest site?

We propose three different, non-exclusive mechanisms to explain home-site fidelity. First, honeybees prefer old, abandoned nest sites to new ones¹¹. After long-range migration, swarms initially choose highly attractive bee trees. Scouts then recognize the remains of combs — which hold important cues for honeybee nestmate recognition¹² — and use these to choose a new site. Second, nest-site preference could be genetically variable, so that related scouts choose the same locations in subse-

Table 1 Putative genotypes of queens and their drone mates

Site (year)	A1 (1995)			A1 (1997)			A2 (1995)			A2 (1997)		
Locus	A14	A76	A88	A14	A76	A88	A14	A76	A88	A14	A76	A88
Queen's alleles	207	201	131	203	203	136	205	201	131	207	201	131
	209	211	131	209	209	143	209	211	131	209	213	136
Drone												
1	203	207	145	203	203	138	205	203	137	207	Q	131
2	203	Q	139	205	203	Q	205	201	137	207	201	142
3	205	201	131	205	201	127	205	Q	135	209	205	130
4	205	201	137	205	207	132	203	211	149	211	213	127
5	209	201	139	205	213	134	213	201	133	Q	201	Q
6	209	203	139	205	211	136	205	201	147	Q	203	130
7	209	203	147	207	209	132	209	201	139	Q	209	138
8	209	207	133	207	201	138	209	Q	135	Q	213	134
9	209	211	141	Q	209	ND	205	Q	127			
10	209	213	127	Q	211	132	Q	207	129			
11	Q	203	133				209	Q	139			
12							Q	213	141			
13							209	213	127			

Site (year)	B1 (1995)			B1 (1997)			B2 (1995)			B2 (1997)			B3 (1995)			B3 (1997)		
Locus	A14	A76	A88	A14	A76	A88	A14	A76	A88	A14	A76	A88	A14	A76	A88	A14	A76	A88
Queen's alleles	203	201	131	203	203	136	205	201	127	205	201	127	207	201	129	209	201	129
	209	201	135	207	205	137	209	212	137	209	212	137	209	203	139	211	201	134
Drone																		
1	203	201	139	203	ND	127	205	201	137	205	201	137	203	211	135	203	201	127
2	203	211	133	207	203	144	210	199	Q	210	199	Q	203	213	135	203	213	129
3	203	211	137	207	ND	138	Q	201	135	Q	201	135	205	199	135	205	211	130
4	203	211	139	209	205	138	Q	203	129	Q	203	129	207	201	137	205	199	129
5	205	201	135	209	Q	137	203	199	127	203	199	148	207	Q	139	205	199	171
6	205	201	137	209	201	127	203	201	139	205	213	145	209	201	129	207	201	129
7	205	207	135	209	201	134	203	209	127	207	201	130	209	201	137	207	205	149
8	207	205	Q	209	211	149	203	217	143	211	ND	145	209	203	135	209	205	127
9	207	207	133	Q	213	134	203	Q	139	Q	201	134	209	211	133	209	207	136
10	207	207	135				207	201	139	Q	ND	148	209	211	135	209	ND	138
11	209	201	139				207	205	135	Q	Q	Q	209	213	139	211	205	136
12	209	201	143				209	199	131	ND	ND	145	215	201	133			
13	211	201	133				209	201	133				215	203	127			
14	211	213	Q				209	201	143				215	207	139			
15	217	201	137				209	203	142				Q	201	127			
16	Q	201	133				209	205	129				Q	205	ND			
17	Q	205	137				209	207	133				Q	213	135			
18							209	211	133									
19							209	211	135									
20							203	199	139									
21							203	213	135									
22							215	201	127									

Allele lengths in base pairs for three loci of the queens and their drone mates are shown. Samples from the same nesting location are shown side by side. Bold, matching genotypes; ND, no amplification product; Q, drone has one of the queen's two alleles.

quent years. Third, if there is genetic variance in the onset of swarming, colonies arriving first at bee trees would select the most attractive sites, and so choose the same location as previous years.

Irrespective of which mechanism governs the orientation of migrating swarms, it is subject to genetic variance, allowing genes — but not necessarily individuals — to return to an old nest site.

Peter Neumann*†, Nikolaus Koeniger‡, Gudrun Koeniger‡, Salim Tingek\$, Per Kryger||, Robin F. A. Moritz*

*Institut für Zoologie, Martin-Luther-Universität Halle-Wittenberg, 06099 Halle/Saale, Germany

†Department of Zoology and Entomology, Rhodes University, 6140 Grahamstown, South Africa
e-mail: p.neumann@ru.ac.za

‡Institut für Bienenkunde (Polytechnische Gesellschaft), J. W. Goethe-Universität Frankfurt am Main, 61440 Oberursel, Germany

\$Agricultural Research Station, PO Box 197, 89908 Tenom, Sabah, Malaysia

||Department of Zoology and Entomology, University of Pretoria, 0001 Pretoria, South Africa

1. Koeniger, N. & Koeniger, G. J. *Apic. Res.* **19**, 21–34 (1980).
2. Ruttner, F. *Taxonomy and Biogeography of Honeybees* (Springer, Berlin, 1988).
3. Lindauer, M. Z. *Vergl. Physiol.* **38**, 521–557 (1956).
4. Moritz, R. F. A. *et al. Behav. Ecol. Sociobiol.* **37**, 357–363 (1995).
5. Estoup, A., Solignac, M., Harry, M. & Cornuet, J. M. *Nucleic Acids Res.* **21**, 1427–1431 (1993).
6. Estoup, A., Solignac, M. & Cornuet, J. M. *Proc. R. Soc. Lond. B* **258**, 1–7 (1994).
7. Oldroyd, B. P. *et al. Ann. Entomol. Soc. Am.* **89**, 276–283 (1996).
8. Neumann, P., Moritz, R. F. A. & Mautz, D. *Apidologie* **30**, 505–512 (1999).
9. Thapa, R., Wongsiri, S. & Oldroyd, B. in *Social Insects at the Turn of the Millennium* (eds Schwarz, M. P. & Hogendoorn, K.) 469 (IUSI, Adelaide, 1999).
10. Otis, G. W., Mardan, M. & McGee, K. in *Social Insects and the Environment* (eds Veeresh, G. K., Mallik, B. & Viraktamath, C. A.) 378 (Oxford & IBH, New Delhi, 1990).
11. Visscher, P. K., Morse, R. A. & Seeley, T. D. *Ins. Soc.* **32**, 217–220 (1985).
12. Breed, M. D. *et al. Anim. Behav.* **50**, 489–496 (1995).

Entomology

Giant honeybees return to their nest sites

The Asian giant honeybee *Apis dorsata* forms massive single-comb colonies which usually hang from a tree branch or the eaves of buildings. Although colonies regularly migrate over many kilometres, we find that they often return to their original nest site — even after an absence of up to two years. How the bees do this is unknown, as workers live for only a few weeks.

A. dorsata colonies are often found in dense aggregations: there can be more than 200 on a single tree. Their migration pattern is seasonal and may be related to food availability or predation. After occupying a nest site for several months, all or most colonies abscond and travel to an alternative site, which may be 200 km away¹.

Mangaldai					Khulsi										
Season	Locus	Queen genotype			Season	Locus	Queen genotype								
1997–98	A88	A	B	C	1996–97	A88	F	G	H	I	J				
	A14	137/142	132/134	134/137		A88	137/143	136/136	141/141	143/147	138/140				
	A76	206/212	206/206	212/212		A14	206/212	205/209	214/214	202/214	206/206				
	A107	204/208	206/208	204/210		A76	204/210	206/212	202/206	202/214	204/204				
1998–99	A107	215/245	171/189	220/245	1998–99	A107	218/251	170/195	195/243	218/243	218/247				
	A88	D	E	1999–2000		A88	K	L	M	1999–2000	A88	N	O	P	Q
	A14	137/142	137/147			A88	137/143	140/145	141/150		A88	137/143	139/141	141/150	137/137
	A76	206/212	209/213			A14	206/212	205/211	207/212		A14	206/212	207/207	207/212	205/309
A107	204/208	204/206	A76		204/210	208/216	204/208	A76	204/210		208/212	204/208	206/206		
1999–2000	A107	215/245	210/245	A107	218/251	215/246	218/235	A107	218/251	210/252	218/235	210/239			

Figure 1 Inferred genotypes of queens (A–Q) heading colonies of *Apis dorsata* at two sites in India. Colonies with the same shading are almost certainly headed by the same queen. The probability that different queens could have the same genotype by chance alone is approximately 2.24×10^{-5} for colonies A and D, 6.06×10^{-8} for colonies F, K and N, and 5.28×10^{-7} for colonies M and P. To calculate these probabilities, we estimated the frequency of each allele among the drones that fathered the sampled workers pooled over the three years of the study ($n = 150$ paternal alleles at Mangaldai, 309 at Khulsi). For a locus of genotype AB, the probability that any two queens would have the same genotype by chance is $2pq$, where p is the frequency of allele A in the population, and q is the frequency of allele B. The overall probability that two random queens would share the same genotype is the product of the individual probabilities for each locus (assuming no linkage between loci). These probabilities are biased to an unknown degree because, first, unequal sampling of spermatozoa during fertilization may mean that the array of paternal alleles among the sampled workers deviated significantly from the overall population and, second, the probability may be understated owing to consanguineous matings (although these are thought to be rare in *Apis*). However, the probability that any two queens would have the genotypes observed by chance alone is still extremely remote.

Indigenous honey hunters have long known that migrating swarms occupy certain nest sites year after year, and we know of one tree in Malaysia that has been repeatedly occupied by about 100 colonies since 1965, and possibly much earlier².

After the colonies leave their nesting site, they are usually absent for several months. To test whether swarms return to their natal nesting site, we measured the relatedness of colonies at two sites in Assam, India, between 1996 and 2000. Both sites were houses: one was in the town of Mangaldai (26° 26' N, 92° 02' E) and the other was in the village of Khulsi (26° 44' N, 91° 33' E).

The owners of the house in Mangaldai told us that three nests have occupied it seasonally for the past 15 years. We obtained samples in 1997–99. The house in Khulsi is abandoned, and was occupied by 3–10 nests. We first obtained samples in 1996–97, but in 1997–98 bees were excluded from the house by human occupation. These people left in 1998, and we obtained samples in 1998–99 and 1999–2000. We determined the genotype of each colony's queen by microsatellite-based pedigree analysis using the primers A14, A76, A88 and A107 (refs 3,4).

The probability that any two queens sampled at random will share the same genotype by chance alone is given by the product rule of forensic genetics⁵. Of the 17 colonies studied, 4 pairs were almost certainly headed by the same queen (Fig. 1). This means that these queens migrated to

new nest sites, and returned the following season. Queen F returned twice: as K in 1998–99, and as N in 1999–2000. This queen could not return to Khulsi in 1997–98 because people occupied the house.

These results show that migrating *A. dorsata* swarms are astonishingly faithful to their nest sites. Returning swarms usually have hundreds of alternative nest sites, but seem to return to the same structure or tree in preference to all others. They may even return to the same eave or branch. It is not known how they do this, but the mechanism is probably related to chemical signatures left by the departing swarm and recognized by the returning swarm.

Although the unavailability of the house in Khulsi in 1997–98 did not deter colony F from returning (as K) in 1998–99, our results demonstrate the need for conserving extant nest sites for *A. dorsata*, an important tropical pollinator and source of food and wax for people in Asia.

J. Paar*, B. P. Oldroyd*, G. Kastberger†

*School of Biological Sciences A12, University of Sydney, New South Wales 2006, Australia
e-mail: boldroyd@bio.usyd.edu.au

†University of Graz, Institute of Zoology, Universitätsplatz 2, A-8010 Graz, Austria

1. Koeniger, N. & Koeniger, G. J. *Apic. Res.* **19**, 21–34 (1980).
2. Oldroyd, B. P., Osborne, K. E. & Mardan, M. *Ins. Soc.* **47**, 94–95 (2000).
3. Estoup, A. *et al. Proc. R. Soc. Lond. B* **258**, 1–7 (1994).
4. Oldroyd, B. P. & Osborne, K. E. *Proc. R. Soc. Lond. B* **266**, 1335–1339 (1999).
5. Weir, B. S. *Annu. Rev. Genet.* **28**, 597–621 (1994).

DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*

John F. Heidelberg*, Jonathan A. Eisen*, William C. Nelson*, Rebecca A. Clayton, Michelle L. Gwinn*, Robert J. Dodson*, Daniel H. Haft*, Erin K. Hickey*, Jeremy D. Peterson*, Lowell Umayam*, Steven R. Gill*, Karen E. Nelson*, Timothy D. Read*, Hervé Tettelin*, Delwood Richardson*, Maria D. Ermolaeva*, Jessica Vamathevan*, Steven Bass*, Haiying Qin*, Ioana Dragoi*, Patrick Sellers*, Lisa McDonald*, Teresa Utterback*, Robert D. Fleishmann*, William C. Nierman*, Owen White*, Steven L. Salzberg*, Hamilton O. Smith*†, Rita R. Colwell‡, John J. Mekalanos§, J. Craig Venter*† & Claire M. Fraser*

* The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA

† Center of Marine Biotechnology, University of Maryland Biotechnology Institute, 701 East Pratt Street, Baltimore, Maryland 21202, USA, and Department of Cell and Molecular Biology, University of Maryland, College Park, Maryland 20742, USA

§ Harvard Medical School, Department of Microbiology and Molecular Genetics, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

Here we determine the complete genomic sequence of the Gram negative, γ -Proteobacterium *Vibrio cholerae* El Tor N16961 to be 4,033,460 base pairs (bp). The genome consists of two circular chromosomes of 2,961,146 bp and 1,072,314 bp that together encode 3,885 open reading frames. The vast majority of recognizable genes for essential cell functions (such as DNA replication, transcription, translation and cell-wall biosynthesis) and pathogenicity (for example, toxins, surface antigens and adhesins) are located on the large chromosome. In contrast, the small chromosome contains a larger fraction (59%) of hypothetical genes compared with the large chromosome (42%), and also contains many more genes that appear to have origins other than the γ -Proteobacteria. The small chromosome also carries a gene capture system (the integron island) and host 'addiction' genes that are typically found on plasmids; thus, the small chromosome may have originally been a megaplasmid that was captured by an ancestral *Vibrio* species. The *V. cholerae* genomic sequence provides a starting point for understanding how a free-living, environmental organism emerged to become a significant human bacterial pathogen.

Vibrio cholerae is the aetiological agent of cholera, a severe diarrhoeal disease that occurs most frequently in epidemic form¹. Cholera has been epidemic in southern Asia for at least 1,000 years, but also spread worldwide to cause seven pandemics since 1817 (ref. 1). When untreated, cholera is a disease of extraordinarily rapid onset and potentially high lethality. Although clinical management of cholera has advanced over the past 40 years, cholera remains a serious threat in developing countries where sanitation is poor, health care limited, and drinking water unsafe.

Vibrio cholerae as a species includes both pathogenic and nonpathogenic strains that vary in their virulence gene content². This bacterium contains a wide variety of strains and biotypes, receiving and transferring genes for toxins³, colonization factors^{4,5}, antibiotic resistance⁶, capsular polysaccharides that provide resistance to chlorine⁷ and new surface antigens, such as the O139 lipopolysaccharide and O antigen capsule^{8,9}. The lateral or horizontal transfer of these virulence genes by phage³, pathogenicity islands^{10,11} and other accessory genetic elements¹² provides insights into how bacterial pathogens emerge and evolve to become new strains.

Vibrio species represent a significant portion of the culturable heterotrophic bacteria of oceans, coastal waters and estuaries^{13,14}. Environmental studies show that these bacteria strongly influence nutrient cycling in the marine environment. Various species of this genus are also devastating pathogens for finfish, shellfish and mammals. There is still much to be learned about the aquatic ecology and natural history of *V. cholerae* including its autochthonous (native) existence in endemic locales during cholera-free, interepidemic periods, which environmental factors, such as climate^{13,15}, aided its re-emergence in Latin America, and which environmental factors are associated with its habitat in cholera endemic regions. For example, *V. cholerae*, during interepidemic periods, is an inhabitant

of brackish and estuarine waters, and in these environments is associated with zooplankton and other aquatic flora and fauna¹⁶. The organism also enters a "viable but nonculturable"¹⁷ state under certain conditions. Roles for these environmental interactions and this dormant physiological state in the emergence and persistence of pathogenic *V. cholerae* have been proposed^{14,17}.

Here we report the determination and analysis of the *Vibrio cholerae* genome sequence. This analysis represents an important step toward the complete molecular description of how this free-living environmental organism emerged to become a human pathogen by horizontal gene transfer.

Genome analysis

The genome of *V. cholerae* was sequenced by the whole genome random sequencing method¹⁸. The genome consists of two circular chromosomes^{19,20} of 2,961,146 (chromosome 1) and 1,072,314

Table 1 General features of the *Vibrio cholerae* genome

	Chromosome 1	Chromosome 2
Size (bp)	2,961,151	1,072,914
Total number of sequences	36,797	14,367
G+C percentage	47.7	46.9
Total number of ORFs	2,770	1,115
ORF size (bp)	952	918
Percentage coding	88.6	86.3
Number of rRNA operons (16S-23S-5S)	8	0
Number of tRNA	94	4
Number similar to known proteins	1,614 (58%)	465 (42%)
Number similar to proteins of unknown function*	163 (6%)	66 (6%)
Number of conserved hypothetical proteins†	478 (17%)	165 (15%)
Number of hypothetical proteins‡	515 (19%)	419 (38%)
Number of Rho-independent terminators	599	193

bp, base pairs. ORFs, open reading frames.

* Proteins of unknown function, significant sequence similarity (homology) to a named protein for which there is currently no known function.

† Conserved hypothetical protein, sequence similarity to a translation of another ORF, but there is currently no experimental evidence a protein is expressed.

‡ Hypothetical protein, no significant sequence similarity to another protein.

† Present address: Celera Genomics, 45 West Gude Drive, Rockville, Maryland 20850, USA

(chromosome 2) base pairs, with an average G+C content of 46.9% and 47.7%, respectively. There are a total of 3,885 predicted open reading frames (ORFs) and 792 predicted Rho-independent terminators; with 2,770 and 1,115 ORFs and 599 and 193 Rho-independent terminators on the individual chromosomes (Table 1, Figs 1 and 2). Most genes required for growth and viability are located on chromosome 1, although some genes found only on chromosome 2 are also thought to be essential for normal cell function (for example, *dsdA*, *thrS* and the genes encoding ribosomal proteins L20 and L35). Additionally, many intermediaries of metabolic pathways are encoded only on chromosome 2 (Fig. 3).

The replicative origin in chromosome 1 was identified by similarity to the *Vibrio harveyi* and *Escherichia coli* origins, co-localization of genes (*dnaA*, *dnaN*, *recF* and *gyrA*) often found near the origin in prokaryotic genomes, and GC nucleotide skew (G–C/G+C) analysis²¹. Based on these, we designated base-pair 1 in an intergenic region that is located in the putative origin of replication. Only the GC skew analysis was useful in identifying a putative origin on chromosome 2.

This genomic sequence of *V. cholerae* confirmed the presence of a large integron island (a gene capture system) located on chromosome 2 (125.3 kbp)^{22,23}. The *V. cholerae* integron island contains all copies of the *V. cholerae* repeat (VCR) sequence and 216 ORFs (Fig. 1). However, most of these ORFs have no homology to other sequences. Among the recognizable integron island genes are three that encode gene products that may be involved in drug resistance (chloramphenicol acetyltransferase, fosfomycin resistance protein and glutathione transferase), several DNA metabolism enzymes (MutT, transposase, and an integrase), potential virulence genes (haemagglutinin and lipoproteins) and three genes which encode gene products similar to the 'host addiction' proteins (*higA*, *higB* and *doc*), which are used by plasmids to select for their maintenance by host cells.

Comparative genomics

The two-chromosome structure of *V. cholerae* allows for comparisons, both between the two chromosomes of this organism and between either of the *V. cholerae* chromosomes and the chromosomes of other microbial species. There is pronounced asymmetry in the distribution of genes known to be essential for growth and virulence between the two chromosomes. Significantly more genes encoding DNA replication and repair, transcription, translation, cell-wall biosynthesis and a variety of central catabolic and biosynthetic pathways are encoded by chromosome 1. Similarly, most genes known to be essential in bacterial pathogenicity (that is, those encoding the toxin co-regulated pilus, cholera toxin, lipopolysaccharide and the extracellular protein secretion machinery) are also located on chromosome 1. In contrast, chromosome 2 contains a larger fraction (59%) of hypothetical genes and genes of unknown function, compared with chromosome 1 (42%) (Fig. 4). This partitioning of hypothetical proteins on chromosome 2 is highly localized in the integron island (Fig. 2). Chromosome 2 also carries the 3-hydroxy-3-methylglutaryl CoA reductase, a gene apparently acquired from an archaea (Y. Boucher and W. F. Doolittle, personal communication).

The majority of the *V. cholerae* genes were very similar to *E. coli* genes (1,454 ORFs), but 499 (12.8%) of the *V. cholerae* ORFs showed highest similarity to other *V. cholerae* genes, suggesting recent duplications (Figs 5 and 6). Most of the duplicated ORFs encode products involved in regulatory functions (59), chemotaxis (50), transport and binding (42), transposition (18), pathogenicity (13) or unknown functions encoded by conserved hypothetical (62) and hypothetical proteins (113). There are 105 duplications with at least one of each ORF on each chromosome indicating there have been recent crossovers between chromosomes. The extensive duplication of genes involved in scavenging behaviour (chemotaxis and solute transport) suggests the importance of these gene products in

V. cholerae biology, notably its ability to inhabit diverse environments. These environments, in turn, may have selected the duplication and divergence of genes useful for specialized functions. Additionally, whereas El Tor strain N16961 carries only a single copy of the cholera toxin prophage, other *V. cholerae* strains carry several copies of this element^{24,25}, and strains of the classical biotype have a second copy of the prophage that is localized on chromosome 2 (ref. 20). Thus, virulence genes are presumed to be subject to selective pressure, affecting copy number and chromosomal location.

Several ORFs with apparently identical functions exist on both chromosomes which were probably acquired by lateral gene transfer. For example, *glyA* (encoding serine hydroxymethyl transferase) is found once on each chromosome but the phylogenetic analysis suggest the *glyA* copy on chromosome 1 branches with the α -Proteobacteria, whereas the copy on chromosome 2 branches with the γ -Proteobacteria (see Supplementary Information). The chromosome 2 *glyA* is flanked by genes encoding transposases, suggesting that this gene was acquired through a transposition event.

Figure 1 Linear representation of the *V. cholerae* chromosomes. The location of the predicted coding regions, colour-coded by biological role, RNA genes, tRNAs, other RNAs, Rho-independent terminators and *Vibrio cholerae* repeats (VCRs) are indicated. Arrows represent the direction of transcription for each predicted coding region. Numbers next to the tRNAs represent the number of tRNAs at a locus. Numbers next to GES represent the number of membrane-spanning domains predicted by the Goldman, Engleman and Steitz scale calculated by TopPred for that protein. Gene names are available at the TIGR web site (www.tigr.org) and as Supplementary Information.

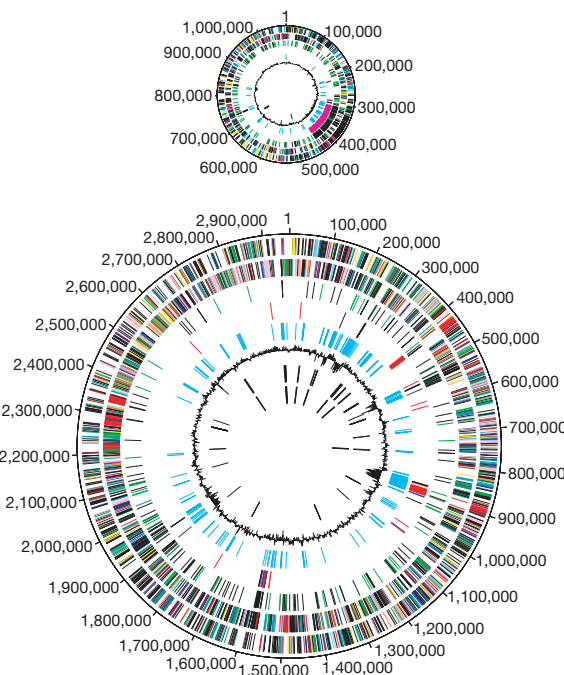


Figure 2 Circular representation of the *V. cholerae* genome. The two chromosomes, large and small, are depicted. From the outside inward: the first and second circles show predicted protein-coding regions on the plus and minus strand, by role, according to the colour code in Fig. 1 (unknown and hypothetical proteins are in black). The third circle shows recently duplicated genes on the same chromosome (black) and on different chromosomes (green). The fourth circle shows transposon-related (black), phage-related (blue), VCRs (pink) and pathogenesis genes (red). The fifth circle shows regions with significant χ^2 values for trinucleotide composition in a 2,000-bp window. The sixth circle shows percentage G+C in relation to mean G+C for the chromosome. The seventh and eighth circles are tRNAs and rRNAs, respectively.

Origin and function of the small chromosome of *V. cholerae*

Several lines of evidence suggest that chromosome 2 was originally a megaplasmid captured by an ancestral *Vibrio* species. The phylogenetic analysis of the ParA homologues located near the putative origin of replication of each chromosome shows chromosome 1 ParA tending to group with other chromosomal ParAs, and the ParA from chromosome 2 tending to group with plasmid, phage and megaplasmid ParAs (see Supplementary Information). In general, genes on chromosome 2, with an apparently identical functioning copy on chromosome 1, appear less similar to orthologues present in other γ -Proteobacteria species (see Supplementary Information). Also, chromosome 1 contains all the ribosomal RNA operons and at least one copy of all the transfer RNAs (four tRNAs are found on chromosome 2, but there are duplicates on chromosome 1). In addition, chromosome 2 carries the integron region, an element often found on plasmids²⁶. Finally, the bias in the functional gene content is more easily explained, if chromosome 2

was originally a megaplasmid (Fig. 4). The megaplasmid presumably acquired genes from diverse bacterial species before its capture by the ancestral *Vibrio*. The relocation of several essential genes from chromosome 1 to the megaplasmid completed the stable capture of this smaller replicon. Apparently this capture of the megaplasmid occurred long enough ago that the trinucleotide composition and percentage G+C content between the two chromosomes have become similar (except for laterally moving elements such as the integron island, bacteriophage genomes, transposons, and so on). The two chromosome structure is found in other *Vibrio* species¹⁹ suggesting that the gene content of the megaplasmid continues to provide *Vibrio* with an evolutionary advantage, perhaps within the aquatic ecosystem where *Vibrio* species are frequently the dominant microorganisms^{14,16}.

It is unclear why chromosome 2 has not been integrated into chromosome 1. Perhaps chromosome 2 plays an important specialized function that provides the evolutionary selective pressure to

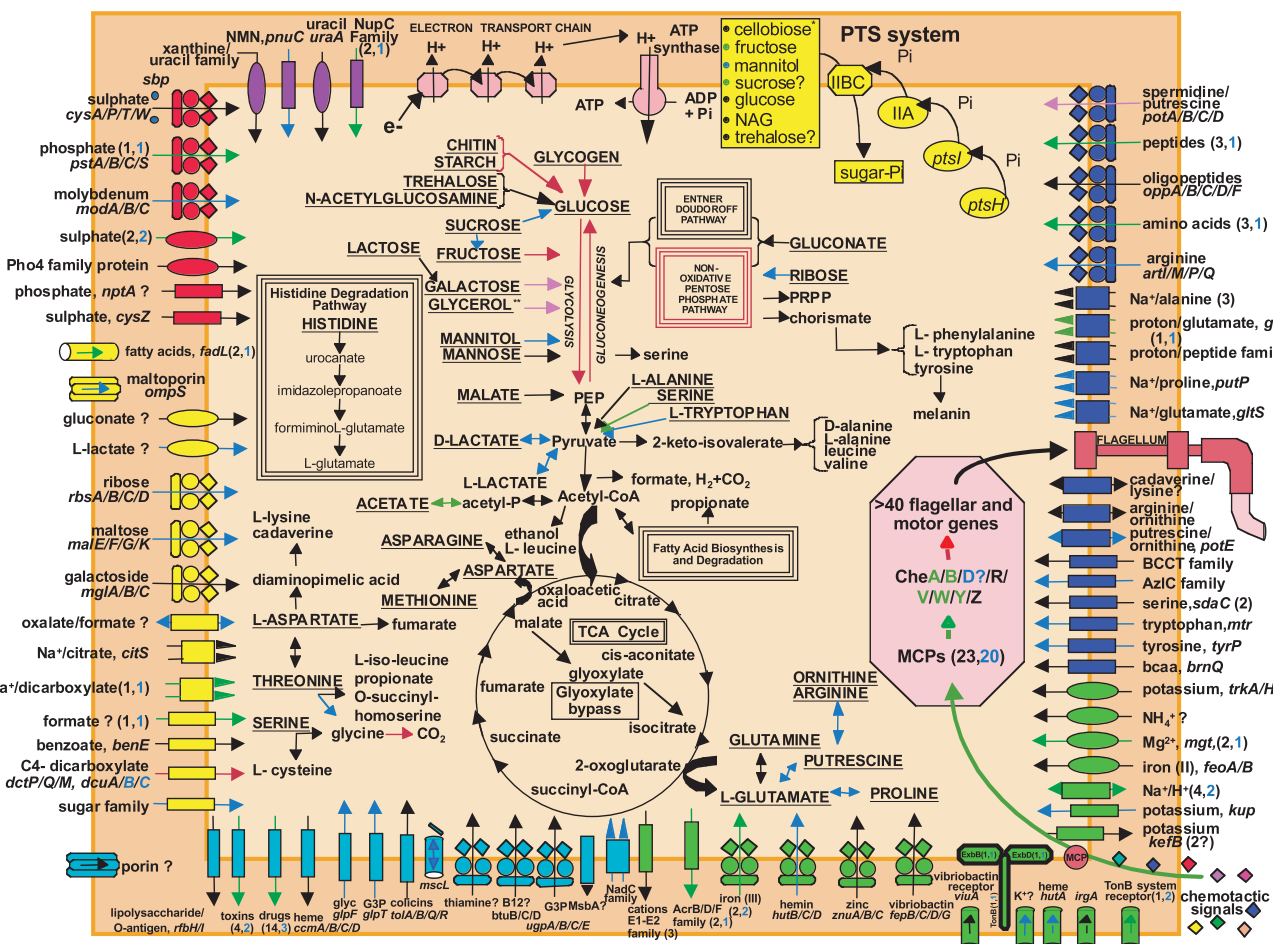


Figure 3 Overview of metabolism and transport in *V. cholerae*. Pathways for energy production and the metabolism of organic compounds, acids and aldehydes are shown. Transporters are grouped by substrate specificity: cations (green), anions (red), carbohydrates (yellow), nucleosides, purines and pyrimidines (purple), amino acids/peptides/amines (dark blue) and other (light blue). Question marks associated with transporters indicate a putative gene, uncertainty in substrate specificity, or direction of transport. Permeases are represented as ovals; ABC transporters are shown as composite figures of ovals, diamonds and circles; porins are represented as three ovals; the large-conductance mechanosensitive channel is shown as a gated cylinder; other cylinders represent outer membrane transporters or receptors; and all other transporters are drawn as rectangles. Export or import of solutes is designated by the direction of the arrow through the transporter. If a precise substrate could not be determined for a transporter, no gene name was assigned and a more general common name reflecting the type of substrate being transported was used. Gene location on the two chromosomes, for both

transporters and metabolic steps, is indicated by arrow colour: all genes located on the large chromosome (black); all genes located on the small chromosome (blue); all genes needed for the complete pathway on one chromosome, but a duplicate copy of one or more genes on the other chromosome (purple); required genes on both chromosomes (red); complete pathway on both chromosomes (green). (Complete pathways, except for glycerol, are found on the large chromosome.) Gene numbers on the two chromosomes are in parentheses and follow the colour scheme for gene location. Substrates underlined and capitalized can be used as energy sources. PRPP, phosphoribosyl-pyrophosphate; PEP, phosphoenolpyruvate; PTS, phosphoenolpyruvate-dependent phosphotransferase system; ATP, adenosine triphosphate; ADP, adenosine diphosphate; MCP, methyl-accepting chemotaxis protein; NAG, N-acetylglucosamine; G3P, glycerol-3-phosphate; glyc, glycerol; NMN, nicotinamide mononucleotide. Asterisk, because *V. cholerae* does not use cellobiose, we expect this PTS system to be involved in chitobiose transport.

suppress integration events when they do occur. For example, if under some environmental condition there is a difference in copy number between the chromosomes, then chromosome 2 may have accumulated genes that are better expressed at higher or lower copy number than genes on chromosome 1. A second possibility is that, in response to environmental cues, one chromosome may partition to daughter cells in the absence of the other chromosome (aberrant segregation). Such single-chromosome-containing cells would be replication-defective but still maintain metabolic activity ('drone' cells), and, therefore, be a potential source of "viable, but non-culturable (VBNC)" cells observed to occur in *V. cholerae*¹⁷. Such 'drone' cells may also play a role in *V. cholerae* biofilms^{7,27,28} by, for example, producing extracellular chitinase, protease and other degradative enzymes that enhance survival of cells in a biofilm, retaining two chromosomes without directly competing with these cells for nutrients.

Transport and energy metabolism

Vibrio cholerae has a diverse natural habitat that includes association with zooplankton in a sessile stage, a planktonic state in the water column, and the capacity to act as a pathogen within the human gastrointestinal tract. It is, therefore, no surprise that this organism maintains a large repertoire of transport proteins with broad substrate specificity and the corresponding catabolic pathways to enable it to respond efficiently to these different and constantly changing ecosystems (Fig. 4). Many of the sugar transporter systems and their corresponding catabolic pathway enzymes are localized on a single chromosome (that is, ribose and lactate transport and degradation enzymes are contained on chromosome 2, whereas the trehalose systems reside on chromosome 1). However, many of the other energy metabolism pathways are split (that is, chitin, glycolysis, and so on) between the chromosomes (Fig. 3).

In aquatic environments, chitin often represents a source of both carbon and nitrogen. This energy source is important for *V. cholerae* as it is associated with zooplankton, which have a chitinous exoskeleton^{13,15,29}. *Vibrio cholerae* degrades chitin by a pathway that is very similar to that of *Vibrio furnissii*³⁰. Sequence analysis suggests a phosphoenolpyruvate phosphotransferase system (PTS)

for cellobiose transport, but as *V. cholerae* does not use cellobiose, it is more likely that this PTS is involved in transport of the structurally similar compound, chitobiose, analogous to the situation proposed for *Bourrelia burgdorferi*¹⁸.

The three anions that are transported by ABC transport systems in *V. cholerae* are molybdenum, phosphate and sulphate. Molybdenum transport genes (*modA/B/C*) are all located on chromosome 2, and most of the sulphate transport genes are on the large molecule. However, copies of the genes for phosphate transporters are found in both chromosomes. The genes in these two phosphate transport operons are different from each other and do not represent a recent duplication; instead, this suggests that one may be an acquired operon.

Interchromosomal regulation

Several of the regulatory pathways, both for regulation in response to environmental and pathogenic signals, are divided between the two chromosomes. These included pathways for starvation survival, 'quorum sensing' and expression of the enterotoxigenic haemolysin, HlyA.

During periods of nutrient starvation, *V. cholerae*, and other Gram-negative bacteria, enter the stationary phase and, later, the viable but nonculturable (VBNC) state^{14,17}. The alternative sigma factor σ^{38} (*rpoS*) is required for survival of *V. cholerae* in the environment but not for pathogenicity³¹, and therefore probably plays an important role in the initiation of the VBNC state. There is one copy of *rpoS*, located on chromosome 1, near the *oriC*. The RpoS regulates expression of several other proteins, including catalase, cyclopropane-fatty-acyl-phospholipid synthase and HA/protease, which are found on both chromosomes³¹.

Genes involved in 'quorum sensing', or cell-density-dependent regulation, also exist on both chromosomes of *V. cholerae*. In bioluminescent *Vibrio* species (notably *Vibrio fischeri* and *V. harveyi*), quorum sensing is used to control light production. Although this strain of *V. cholerae* lacks the genes for bioluminescence, it does have the genes required for the autoinducer-2 (AI-2) quorum-sensing mechanism³² (*luxOPQS*) but this pathway is split between the chromosomes with *luxOSU* on chromosome 1 and *luxPQ* on chromosome 2. Similarly, another transcriptional regulatory gene,

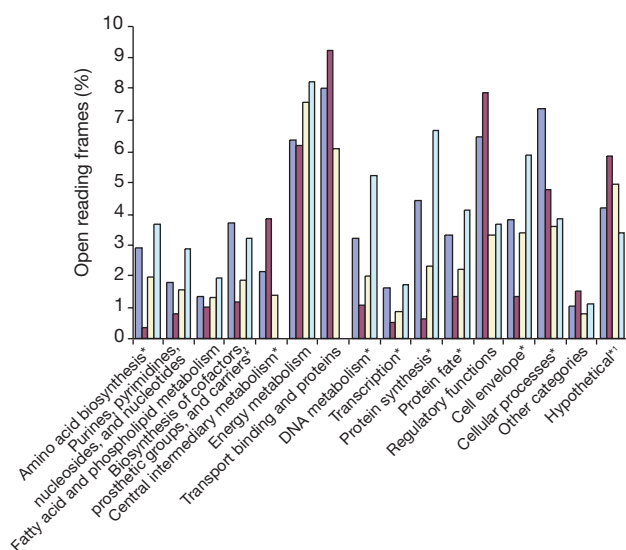


Figure 4 Percentage of total *Vibrio cholerae* open reading frames (ORFs) in biological roles compared with other γ -Proteobacteria. These were *V. cholerae*, chromosome 1 (blue); *V. cholerae*, chromosome 2 (red); *Escherichia coli* (yellow); *Haemophilus influenzae* (pale blue). Significant partitioning ($P < 0.01$) of biological roles between *V. cholerae* chromosomes is indicated with an asterisk, as determined with a χ^2 analysis. 1, Hypothetical contains both conserved hypothetical proteins and hypothetical proteins, and is at 1/10 scale compared with other roles.

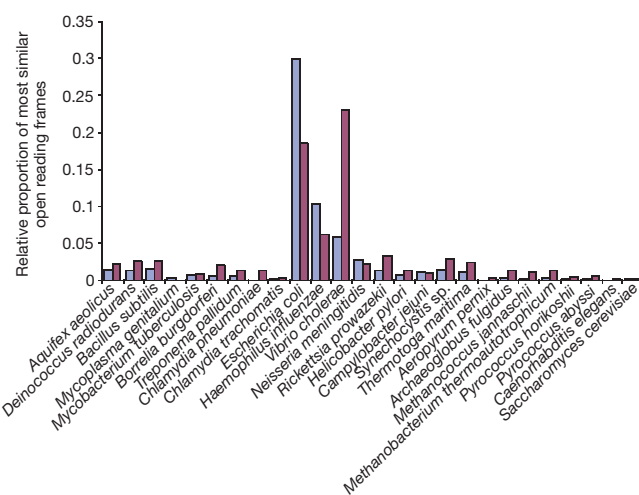


Figure 5 Comparison of the *V. cholerae* ORFs with those of other completely sequenced genomes. The sequence of all proteins from each completed genome were retrieved from NCBI, TIGR and the *Caenorhabditis elegans* (wormpep16) databases. All *V. cholerae* ORFs (large chromosome, blue; small chromosome, red) were searched against all other genomes with FASTA3. The number of *V. cholerae* ORFs with greatest similarity ($E \leq 10^{-5}$) are shown in proportion to the total number of ORFs in that genome. There were no ORFs that were most similar to a *Mycoplasm pneumoniae* ORF.

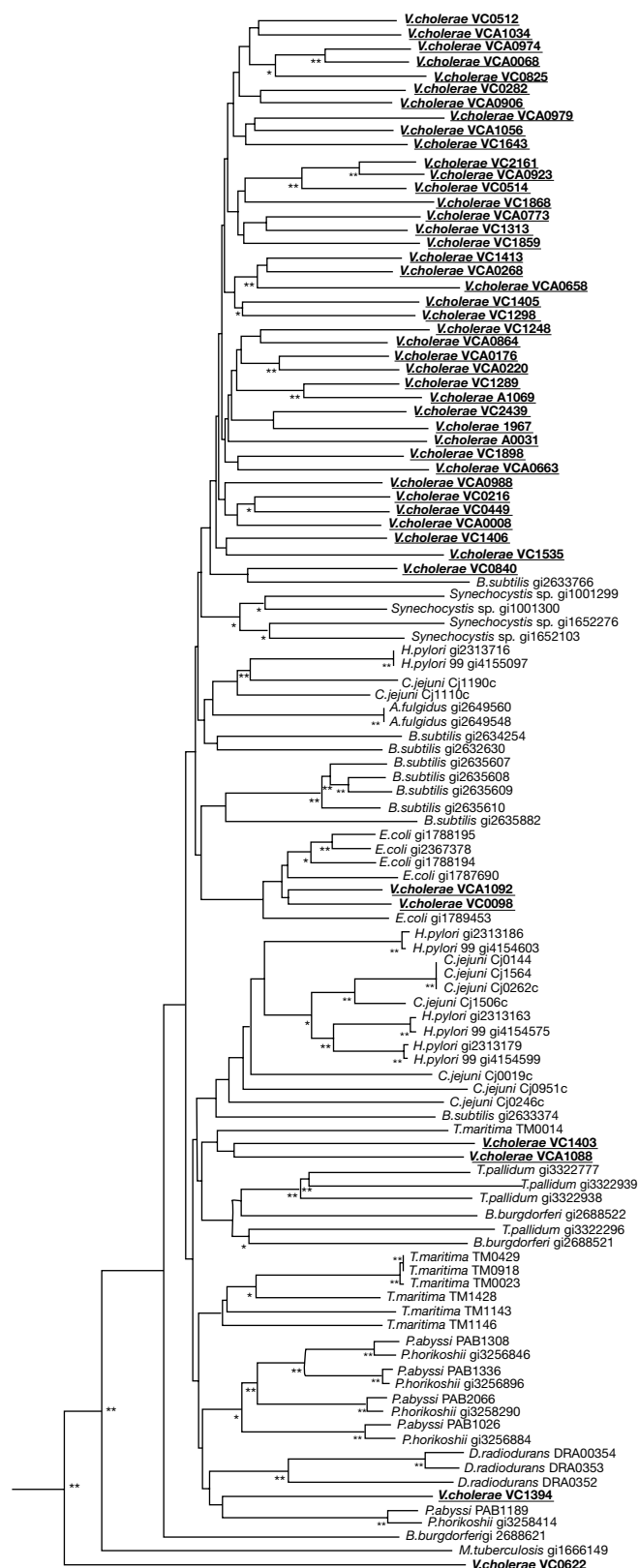


Figure 6 Phylogenetic tree of methyl-accepting chemotactic proteins (MCP) homologues in completed genomes. Homologues of MCP were identified by FASTA3 searches of all available complete genomes. Amino-acid sequences of the proteins were aligned using CLUSTALW, and a neighbour-joining phylogenetic tree was generated from the alignment using the PAUP* program (using a PAM-based distance calculation). Hypervariable regions of the alignment and positions with gaps in many of the sequences were excluded from the analysis. Nodes with significant bootstrap values are indicated: two asterisks, >70%; asterisk, 40–70%.

hlyU (ref. 33), is located on chromosome 1, while the gene it regulates, *hlyA*, is located on chromosome 2.

DNA repair

Vibrio cholerae has genes encoding several DNA-repair and DNA-damage-response pathways, including nucleotide-excision repair, mismatch-excision repair, base-excision repair, AP endonuclease, alkylation transfer, photoreactivation, DNA ligation, and all the major components of recombination and recombinational repair, including initiation, recombination and resolution³⁴. In addition, homologues of many of the genes involved in the SOS response in *E. coli* are found. The presence of three photolyase homologues, more than have been found in other bacterial species, probably allows for the ability to photoreactivate the two major forms of ultraviolet-induced DNA damage (cyclobutane pyrimidine dimers and 6-4 photoproducts), and may also allow use of a range of wavelengths of light used for the energy required for photoreactivation. It is also of interest that many of the repair genes are on chromosome 2 (*alkA*, *ada1*, *ada2*, *phr3*, *mutK*, *sbcCD*, *dcm*, *mutT3*), indicating that this chromosome is probably required for full DNA repair capability.

Pathogenicity

Toxins. The genome sequence of *V. cholerae* El Tor N16961 revealed a single copy of the cholera toxin (CT) genes, *ctxAB*, located on chromosome 1 within the integrated genome of CTX ϕ , a temperate filamentous phage³. The receptor for entry of CTX ϕ into the cell is the toxin-coregulated pilus (TCP)³, and the TCP gene cluster (see below) also resides on chromosome 1. Like the structural genes for CT and TCP, the regulatory gene, *toxR*, which controls their expression *in vivo*³⁵, is also located on chromosome 1.

On the other side of CTX ϕ prophage is a region encoding an RTX toxin (*rtxA*), and its activator (*rtxC*) and transporters (*rtxBD*)³⁶. A third transporter gene has been identified that is a paralogue of *rtxB*, and is transcribed in the same direction as *rtxBD*. Downstream of this gene are two genes encoding a sensor histidine kinase and response regulator. Trinucleotide composition analysis suggests that the RTX region was horizontally acquired along with the sensor histidine kinase/response regulator, suggesting these regulators effect expression of the closely linked RTX transcriptional units.

Also present are genes encoding numerous potential toxins, including several haemolysins, proteases and lipases. These include *hap*, the haemagglutinin protease, a secreted metalloprotease that seems to attack proteins involved in maintaining the integrity of epithelial cell tight junctions³⁷, and *hlyA*, encoding a secreted haemolysin that displays enterotoxic activity³⁸. In contrast to CTX, RTX and all known intestinal colonization factors, the *hap* and *hlyA* genes virulence factors reside on chromosome 2.

Vibrio cholerae has been reported to produce shiga-like toxins³⁹; however, the sequence did not reveal genes encoding specific homologues of the A or B subunits of shiga toxin. Also not detected were genes encoding homologues of *E. coli* heat stable toxin (ST), which have been detected in other pathogenic strains of *V. cholerae*⁴⁰.

Colonization factors. The critical intestinal colonization factor of *V. cholerae* is the TCP, a type IV pilus^{5,41}. The genome sequence confirmed that the genes involved in TCP assembly (*tcpABCDEFGHIJNQRST*) reside on chromosome 1 (ref. 20) as part of a proposed 'pathogenicity island' (also referred to as VPI) composed of recently acquired DNA that encodes not only TCP, but also other genes associated with the ToxR regulatory cascade, such as *acfABCD*, *toxT*, *aldA* and *tagAB*^{10,11}. Trinucleotide composition analysis suggest that this 45.3-kb segment begins at a 20-bp site upstream of *aldA*, and encompasses a helicase-related protein and a transcriptional activator which both share homology with bacteriophage proteins. At the other end of the segment of atypical trinucleotide composition is a phage family integrase and the

other copy of the 20-bp site, which is presumably the target for integration of the island onto the chromosome^{10,11}. It has been proposed that the TCP/ACF island corresponds to the genome of a filamentous phage that uses TCP pilin as a coat protein⁴. However, other than the three genes encoding phage-related proteins (that is, the helicase, transcriptional activator and integrase) we could find no other genes on the island that encoded products with significant homology to the conserved gene products of other filamentous phages or the structural proteins of nonfilamentous phages.

The maltose-sensitive haemagglutinin (MSHA) is unique to the El Tor biotype of *V. cholerae*. Initially characterized as a haemagglutinin, it was later found to be a type IV pilus^{42,43}. The MSHA biogenesis (MshHIJKLMNEGF) and structural (MshBACD) proteins are all clustered on chromosome 1. There are no apparent integrases or transposases that might define this region as a pathogenicity island or suggest an origin for it other than *V. cholerae*. In support of this conclusion, trinucleotide composition analysis shows that this region has similar composition to the rest of the chromosome, suggesting that if these genes were acquired it was very early in the *Vibrio* phylogenetic history. Recently, several investigators have reported that MSHA is not required for intestinal colonization, nor does it seem to appreciably affect the efficiency of colonization^{44–46}, but instead plays a role in biofilm formation^{27,28}. Accordingly, this pilus may be important for the environmental fitness of *Vibrio* species rather than for pathogenic potential.

The *pilA* region of *V. cholerae* genome apparently encodes a third type IV pilus, although it has not been visualized⁴⁷. This gene cluster includes a gene encoding a prepilin peptidase (PilD) that is required for the efficient processing of protein complexes with type IV prepilin-like signal sequences including TCP, MSHA and EPS^{47,48}. The EPS system of *V. cholerae* encodes a type II secretion system involved in extracellular export of CT and other proteins. The EPS system is encoded by chromosome 1 but, like the MSHA genes, trinucleotide analysis suggests that the EPS genes of *V. cholerae* have not been recently acquired. In contrast, trinucleotide composition analysis suggests that the *pilA* gene cluster was acquired by horizontal transfer. Thus, analysis of the *V. cholerae* genome sequence provides some evidence that older gene clusters, like MSHA and EPS, have become dependent on newly acquired genes such as PilD.

Conclusions

The *Vibrio cholerae* genome sequence provides a new starting point for the study of this organism's environmental and pathobiological characteristics. It will be interesting to determine the gene expression patterns that are unique to its survival and replication during human infection³⁵ as well as in the environment^{13,14,16}. Additionally, the genomic sequence of *V. cholerae* should facilitate the study of this model multi-chromosomal prokaryotic organism. Comparative genomics between several species in the genus *Vibrio* will provide a better understanding of the origin of the new small chromosome and the role that it plays in *Vibrio* biology. The genome sequence may also provide important clues to understanding the metabolic and regulatory networks that link genes on the two chromosomes. Finally, *V. cholerae* clearly represents a promising genetic system for studying how several horizontally acquired loci located on separate chromosomes can still efficiently interact at the regulatory, cell biology and biochemical levels. □

Methods

Whole-genome random sequencing procedure.

Vibrio cholerae N16961 was grown from a single isolated colony. Cloning, sequencing and assembly were as described for genomes sequenced by TIGR¹⁸. One small-insert plasmid library (2–3 kb) was generated by random mechanical shearing of genomic DNA. One large insert library was ligated into λ -DASHII/EcoRI vector (Stratagene). In the initial sequence phase, approximately sevenfold sequence coverage was achieved with 49,633 sequences from plasmid clones. Sequences from both ends of 383 λ -clones served as a genome scaffold, verifying the orientation, order and integrity of the contigs. The plasmid and λ sequences were jointly assembled using TIGR Assembler. Sequence gaps were closed

by editing the end sequences and/or primer walking on plasmid clones. Physical gaps were closed by direct sequencing of genomic DNA, or combinatorial polymerase chain reaction (PCR) followed by sequencing the PCR product. The final genome sequence is based on 51,164 sequences.

ORF prediction and gene family identification. An initial set of ORFs, likely to encode proteins, was identified with GLIMMER⁴⁹, and those shorter than 30 codons were eliminated. ORFs that overlapped were visually inspected, and in some cases removed. ORFs were searched against a non-redundant protein database¹⁸. Frameshifts and point mutations were detected and corrected where appropriate. Remaining frameshifts and point mutations are considered to be authentic and were annotated as 'authentic frame-shift' or 'authentic point mutation'. ORFs were also analysed with two sets of hidden Markov models (HMMs) constructed for a number of conserved protein families (1,313 from Pfam v3.1 (ref. 50) and 476 from the TIGRFAM) by use of the HMMER package. TopPred was used to identify membrane-spanning domains in proteins.

Paralogous gene families were constructed by searching the ORFs against themselves using BLASTX, identifying matches with $E \leq 10^{-3}$ over 60% of the query search length, and subsequently clustering these matches into multigene families. Multiple alignments for these protein families were generated with the CLUSTALW program and the alignments scrutinized.

Distribution of all 64 trinucleotides (3-mers) for each chromosome was determined, and the 3-mer distribution in 2,000-bp windows that overlapped by half their length (1,000 bp) across the genome was computed. For each window, we computed the χ^2 statistic on the difference between its 3-mer content and that of the whole chromosome. A large value of this statistic indicates that the 3-mer composition in this window is different from the rest of the chromosome. Probability values for this analysis are based on the assumption that the DNA composition is relatively uniform throughout the genome. Because this assumption may be incorrect, we prefer to interpret high χ^2 values merely as indicators of regions on the chromosome that appear unusual and demand further scrutiny.

Homologues of the genes of interest were identified using the BLASTP and FASTA3 search programs. All homologues were then aligned to each other using the CLUSTALW program with default settings. Phylogenetic trees were generated from the alignments using the neighbour-joining algorithm as implemented by the PAUP* program (with a PAM matrix based distance calculation). Regions of the alignment that were hypervariable or were of low confidence were excluded from the phylogenetic analysis. All alignments are available upon request.

Received 3 April; accepted 18 May 2000.

- Wachsmuth, K., Olsvik, Ø., Evins, G. M. & Popovic, T. in *Vibrio Cholerae And Cholera: Molecular To Global Perspective* (eds Wachsmuth, I. K., Blake, P. A. & Olsvik, Ø.) 357–370 (ASM Press, Washington DC, 1994).
- Faruque, S. M., Albert, M. J. & Mekalanos, J. J. Epidemiology, genetics, and ecology of toxigenic *Vibrio cholerae*. *Microbiol. Mol. Biol. Rev.* **62**, 1301–1314 (1998).
- Waldor, M. K. & Mekalanos, J. J. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* **272**, 1910–1914 (1996).
- Karaolis, D. K., Somara, S., Maneval, D. R. Jr, Johnson, J. A. & Kaper, J. B. A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria. *Nature* **399**, 375–379 (1999).
- Brown, R. C. & Taylor, R. K. Organization of tcp, acf, and toxT genes within a ToxT-dependent operon. *Mol. Microbiol.* **16**, 425–439 (1995).
- Hochhut, B. & Waldor, M. K. Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfC*. *Mol. Microbiol.* **32**, 99–110 (1999).
- Yildiz, F. H. & Schoolnik, G. K. *Vibrio cholerae* O1 El Tor: identification of a gene cluster required for the rugose colony type, exopolysaccharide production, chlorine resistance, and biofilm formation. *Proc. Natl Acad. Sci. USA* **96**, 4028–4033 (1999).
- Bik, E. M., Bunschoten, A. E., Gouw, R. D. & Mooi, F. R. Genesis of the novel epidemic *Vibrio cholerae* O139 strain: evidence for horizontal transfer of genes involved in polysaccharide synthesis. *EMBO J* **14**, 209–216 (1995).
- Waldor, M. K., Colwell, R. & Mekalanos, J. J. The *Vibrio cholerae* O139 serogroup antigen includes an O-antigen capsule and lipopolysaccharide virulence determinants. *Proc. Natl Acad. Sci. USA* **91**, 11388–11392 (1994).
- Kovach, M. E., Shaffer, M. D. & Peterson, K. M. A putative integrase gene defines the distal end of a large cluster of ToxR-regulated colonization genes in *Vibrio cholerae*. *Microbiology* **142**, 2165–2174 (1996).
- Karaolis, D. K. et al. A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains. *Proc. Natl Acad. Sci. USA* **95**, 3134–3139 (1998).
- Mekalanos, J. J., Rubin, E. J. & Waldor, M. K. Cholera: molecular basis for emergence and pathogenesis. *FEMS Immunol. Med. Microbiol.* **18**, 241–248 (1997).
- Colwell, R. R. Global climate and infectious disease: the cholera paradigm. *Science* **274**, 2025–2031 (1996).
- Colwell, R. R. & Spira, W. M. in *Cholera* (eds Barua, D. & Greenough, W. B. III) 107–127 (Plenum Medical, New York, 1992).
- Lobitz, B. et al. Climate and infectious disease: use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. *Proc. Natl Acad. Sci. USA* **97**, 1438–1443 (2000).
- Colwell, R. R. & Huq, A. Environmental reservoir of *Vibrio cholerae*. The causative agent of cholera. *Ann. NY Acad. Sci.* **740**, 44–54 (1994).
- Rozsak, D. B. & Colwell, R. R. Survival strategies of bacteria in the natural environment. *Microbiol. Rev.* **51**, 365–379 (1987).
- Fraser, C. M. et al. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390**, 580–586 (1997).
- Yamaichi, Y., Iida, T., Park, K. S., Yamamoto, K. & Honda, T. Physical and genetic map of the genome of *Vibrio parahaemolyticus*: presence of two chromosomes in *Vibrio* species. *Mol. Microbiol.* **31**, 1513–1521 (1999).
- Trucksis, M., Michalski, J., Deng, Y. K. & Kaper, J. B. The *Vibrio cholerae* genome contains two unique circular chromosomes. *Proc. Natl Acad. Sci. USA* **95**, 14464–14469 (1998).

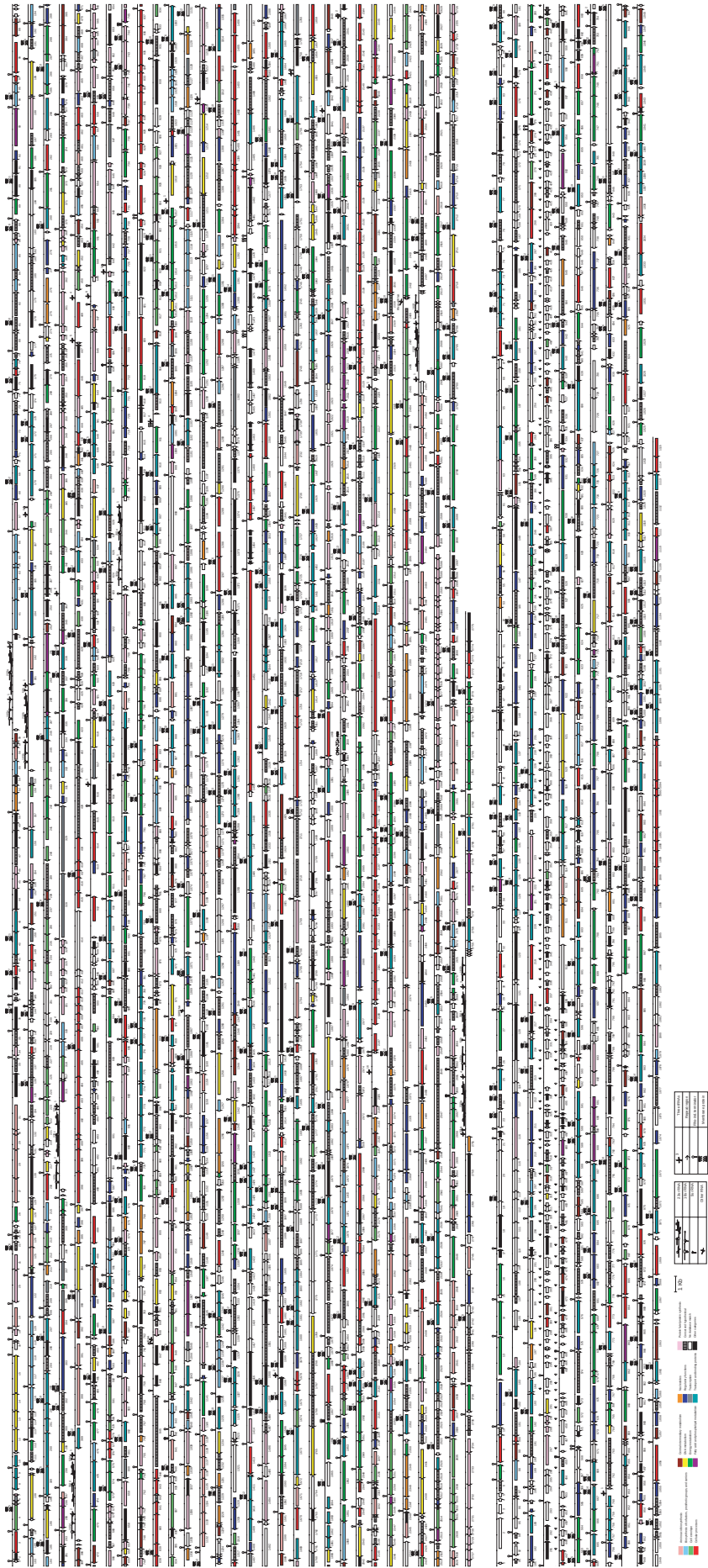
21. Lobry, J. R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13**, 660–665 (1996).
22. Rowe-Magnus, D. A., Guerout, A. M. & Mazel, D. Super-integrans. *Res. Microbiol.* **150**, 641–651 (1999).
23. Hall, R. M., Brookes, D. E. & Stokes, H. W. Site-specific insertion of genes into integrans: role of the 59-base element and determination of the recombination cross-over point. *Mol. Microbiol.* **5**, 1941–1959 (1991).
24. Davis, B. M., Kimsey, H. H., Chang, W. & Waldor, M. K. The *Vibrio cholerae* O139 Calcutta bacteriophage CTXphi is infectious and encodes a novel repressor. *J. Bacteriol.* **181**, 6779–6787 (1999).
25. Mekalanos, J. J. Duplication and amplification of toxin genes in *Vibrio cholerae*. *Cell* **35**, 253–263 (1983).
26. Mazel, D., Dychinco, B., Webb, V. A. & Davies, J. A distinctive class of integron in the *Vibrio cholerae* genome. *Science* **280**, 605–608 (1998).
27. Watnick, P. I. & Kolter, R. Steps in the development of a *Vibrio cholerae* El Tor biofilm. *Mol. Microbiol.* **34**, 586–595 (1999).
28. Watnick, P. I., Fullner, K. J. & Kolter, R. A role for the mannose-sensitive hemagglutinin in biofilm formation by *Vibrio cholerae* El Tor. *J. Bacteriol.* **181**, 3606–3609 (1999).
29. Huq, A. *et al.* Ecological relationships between *Vibrio cholerae* and planktonic crustacean copepods. *Appl. Environ. Microbiol.* **45**, 275–283 (1983).
30. Bassler, B. L., Yu, C., Lee, Y. C. & Roseman, S. Chitin utilization by marine bacteria. Degradation and catabolism of chitin oligosaccharides by *Vibrio furnissii*. *J. Biol. Chem.* **266**, 24276–24286 (1991).
31. Yildiz, F. H. & Schoolnik, G. K. Role of rpoS in stress survival and virulence of *Vibrio cholerae*. *J. Bacteriol.* **180**, 773–784 (1998).
32. Bassler, B. L., Greenberg, E. P. & Stevens, A. M. Cross-species induction of luminescence in the quorum-sensing bacterium *Vibrio harveyi*. *J. Bacteriol.* **179**, 4043–4045 (1997).
33. Williams, S. G., Attridge, S. R. & Manning, P. A. The transcriptional activator HlyU of *Vibrio cholerae*: nucleotide sequence and role in virulence gene expression. *Mol. Microbiol.* **9**, 751–760 (1993).
34. Eisen, J. A. & Hanawalt, P. C. A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat. Res.* **435**, 171–213 (1999).
35. Lee, S. H., Hava, D. L., Waldor, M. K. & Camilli, A. Regulation and temporal expression patterns of *Vibrio cholerae* virulence genes during infection. *Cell* **99**, 625–634 (1999).
36. Lin, W. *et al.* Identification of a *Vibrio cholerae* RTX toxin gene cluster that is tightly linked to the cholera toxin prophage. *Proc. Natl Acad. Sci. USA* **96**, 1071–1076 (1999).
37. Wu, Z., Milton, D., Nybom, P., Sjo, A. & Magnusson, K. E. *Vibrio cholerae* hemagglutinin/protease (HA/protease) causes morphological changes in cultured epithelial cells and perturbs their paracellular barrier function. *Microb. Pathog.* **21**, 111–123 (1996).
38. Alm, R. A., Strocher, U. H. & Manning, P. A. Extracellular proteins of *Vibrio cholerae*: nucleotide sequence of the structural gene (hlyA) for the haemolysin of the haemolytic El Tor strain 017 and characterization of the hlyA mutation in the non-haemolytic classical strain 569B. *Mol. Microbiol.* **2**, 481–488 (1988).
39. O'Brien, A. D., Chen, M. E., Holmes, R. K., Kaper, J. & Levine, M. M. Environmental and human isolates of *Vibrio cholerae* and *Vibrio parahaemolyticus* produce a Shigella dysenteriae 1 (Shiga)-like cytotoxin. *Lancet* **1**, 77–78 (1984).
40. Ogawa, A., Kato, J., Watanabe, H., Nair, B. G. & Takeda, T. Cloning and nucleotide sequence of a heat-stable enterotoxin gene from *Vibrio cholerae* non-O1 isolated from a patient with traveler's diarrhea. *Infect. Immun.* **58**, 3325–3329 (1990).
41. Manning, P. A. The tcp gene cluster of *Vibrio cholerae*. *Gene* **192**, 63–70 (1997).
42. Jonson, G., Holmgren, J. & Svennerholm, A. M. Identification of a mannose-binding pilus on *Vibrio cholerae* El Tor. *Microb. Pathog.* **11**, 433–441 (1991).
43. Jonson, G., Lebens, M. & Holmgren, J. Cloning and sequencing of *Vibrio cholerae* mannose-sensitive haemagglutinin pilin gene: localization of mshA within a cluster of type 4 pilin genes. *Mol. Microbiol.* **13**, 109–118 (1994).
44. Attridge, S. R., Manning, P. A., Holmgren, J. & Jonson, G. Relative significance of mannose-sensitive haemagglutinin and toxin-coregulated pili in colonization of infant mice by *Vibrio cholerae* El Tor. *Infect. Immun.* **64**, 3369–3373 (1996).
45. Tacket, C. O. *et al.* Investigation of the roles of toxin-coregulated pili and mannose-sensitive haemagglutinin pili in the pathogenesis of *Vibrio cholerae* O139 infection. *Infect. Immun.* **66**, 692–695 (1998).
46. Thelin, K. H. & Taylor, R. K. Toxin-coregulated pilus, but not mannose-sensitive haemagglutinin, is required for colonization by *Vibrio cholerae* O1 El Tor biotype and O139 strains. *Infect. Immun.* **64**, 2853–2856 (1996).
47. Fullner, K. J. & Mekalanos, J. J. Genetic characterization of a new type IV-A pilus gene cluster found in both classical and El Tor biotypes of *Vibrio cholerae*. *Infect. Immun.* **67**, 1393–1404 (1999).
48. Marsh, J. W. & Taylor, R. K. Identification of the *Vibrio cholerae* type 4 prepilin peptidase required for cholera toxin secretion and pilus formation. *Mol. Microbiol.* **29**, 1481–1492 (1998).
49. Salzberg, S. L., Delcher, A. L., Kasif, S. & White, O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* **26**, 544–548 (1998).
50. Bateman, A. *et al.* Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res.* **27**, 260–262 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by the National Institutes of Health, National Institute of Allergy and Infectious Disease. We thank M. Heaney, V. Sapiro, B. Lee, M. Holmes and B. Vincent for database and software support.

Correspondence and requests for materials should be addressed to C.M.F. (e-mail: gvc@tigr.org). The annotated genome sequence and the gene family alignments are available at (<http://www.tigr.org/tbd/mdb>). The sequences have been deposited in GenBank with accession number AE003852 (chromosome 1) and AE003853 (chromosome 2).



Evidence for free precession in a pulsar

I. H. Stairs, A. G. Lyne & S. L. Shemar

University of Manchester, Jodrell Bank Observatory, Macclesfield, SK11 9DL, UK

Pulsars are rotating neutron stars that produce lighthouse-like beams of radio emission from their magnetic poles. The observed pulse of emission enables their rotation rates to be measured with great precision. For some young pulsars, this provides a means of studying the interior structure of neutron stars. Most pulsars have stable pulse shapes, and slow down steadily (for example, see ref. 20). Here we report the discovery of long-term, highly periodic and correlated variations in both the pulse shape and the rate of slow-down of the pulsar PSR B1828–11. The variations are best described as harmonically related sinusoids, with periods of approximately 1,000, 500 and 250 days, probably resulting from precession of the spin axis caused by an asymmetry in the shape of the pulsar. This is difficult to understand theoretically, because torque-free precession of a solitary pulsar should be damped out by the vortices in its superfluid interior^{1,2}.

PSR B1828–11 (PSR J1830–1059) is a young pulsar which is observed on about 30 occasions each year using the 76-m Lovell radio telescope at Jodrell Bank; the complete data set spans nearly 13 years. The integrated pulse profile is not stable. It consists of a wide component with a central, narrower peak which varies in intensity on timescales of several months. For each observation, we can determine a ‘shape parameter’ S , which describes the pulse shape, as well as a pulse time-of-arrival (TOA).

Pulsar timing analysis consists of fitting a model of the pulsar rotation to the observed TOAs, and examining the residuals, which are the differences between the TOAs and the best-fit model. If the timing model is correct, the residuals should be gaussianly distributed about zero. The timing residuals for PSR B1828–11 are shown in Fig. 1. More than four cycles of an evolving double-peaked pattern with period of about 1,000 days can be seen. In order to study the variations in more detail, the timing residuals, the rotation period P , the period derivative $\dot{P}$ and the shape parameter S have been calculated for the most recent 2,000 days (Fig. 2). In particular,

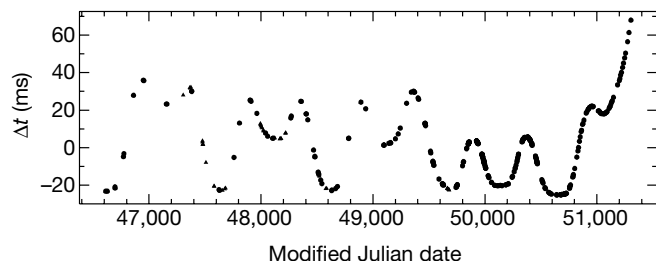


Figure 1 Post-fit timing residuals Δt for PSR B1828–11 after fitting for the spin-down parameters given in Table 1. Each point represents an observation of roughly 30 minutes’ duration at a frequency around 1,400 MHz (circles) or 1,600 MHz (triangles). Individual profiles are described by a linear combination of two extreme standard profiles, one wide, one narrow. An iterative frequency-domain routine was used to determine the relative strengths of the standard profiles required to synthesize the daily profiles. This provides a ‘shape parameter’, $S = \frac{A_N}{A_N + A_W}$, where A_N and A_W are the fitted heights of the narrower and wider standard profiles respectively, so that $S \approx 1$ for the narrowest pulses and $S \approx 0$ for wider ones. The pulse times-of-arrival (TOAs) are also derived from this shape-fitting procedure; the errors in the TOAs are limited by random noise to about 0.2 ms.

Table 1 Measured and derived parameters of PSR B1828–11

Period, P	405.039883158(8) ms
Period derivative, $\dot{P}$	$60.03243(9) \times 10^{-15}$
Period second derivative, $\ddot{P}$	$-1.70(3) \times 10^{-25} \text{ s}^{-1}$
Epoch of period	MJD 48958
Right ascension, $\alpha(\text{J2000})$	18 h 30 min 47.583(7) s
Declination, $\delta(\text{J2000})$	$-10^\circ 59' 29.33(12)''$
Dispersion measure, DM	$159.7(10) \text{ cm}^{-3} \text{ pc}$
Magnetic field, B	$5.0 \times 10^{12} \text{ G}$
Characteristic age, τ_c	0.11 Myr

Uncertainties in least-significant digits are given by numbers in parentheses. Timing parameters were determined using the TEMPO software package (<http://pulsar.princeton.edu/tempo>) and the JPL DE200 planetary ephemeris. The adopted dispersion measure was obtained from same-epoch 1,400-MHz and 600-MHz observations. The pulsar position was determined from timing measurements and refined using interferometric observations with the Very Large Array (VLA). The period first and second derivatives were fitted after whitening the timing residuals by removing a series of high-frequency sinusoids. The choice of the period second derivative, in particular, is therefore somewhat arbitrary, and may affect attempts to model the longest-term variations in the timing residuals.

the variations in pulse shape $\langle S \rangle$ are highly correlated with the changes in the period derivative $\dot{P}$. A spectral analysis of the four time sequences (Fig. 3) shows that all contain harmonically-related periodicities of approximately 1,000, 500 and 250 days.

We consider in turn whether these periodic phenomena originate outside the pulsar in a planetary system, in the neutron superfluid within the star, or in the precession of the star. The harmonically-related period variations alone would suggest a Doppler effect due to a system of gravitationally interacting planets orbiting the pulsar³, as observed for the millisecond pulsar PSR B1257+12 (refs 4–6). However, it does not seem credible that the pulse profile

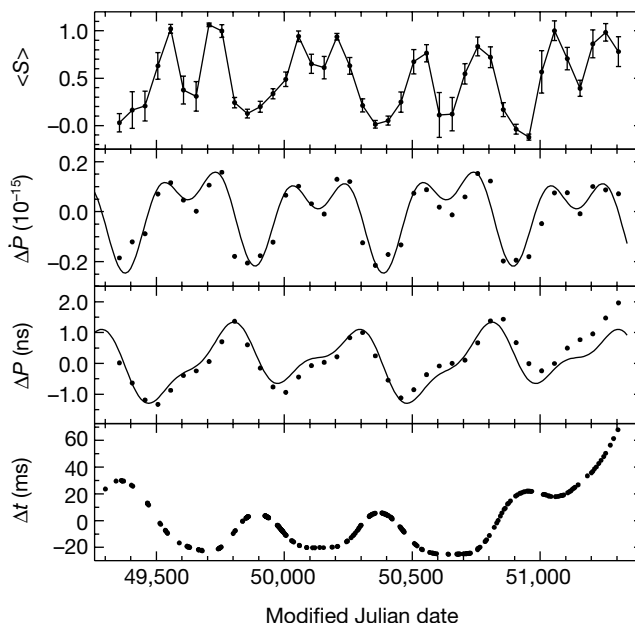


Figure 2 Variations in rotation and pulse shape in PSR B1828–11. The displayed time series are the residuals in arrival time Δt , and period ΔP and derivative $\dot{\Delta P}$, relative to the spin-down model given in Table 1, and the mean pulse shape parameter $\langle S \rangle$ for the most recent 2,000 days, where the observations are most closely spaced. The period residual, ΔP , is the local value of the slope of Δt , and $\dot{\Delta P}$ is the second derivative. The three series ΔP , $\dot{\Delta P}$ and $\langle S \rangle$ were calculated over time intervals of 100 days which overlapped by 50 days. There is, within the uncertainties in $\langle S \rangle$, a clear similarity in form between the variations in $\langle S \rangle$ and in $\dot{\Delta P}$, although this is not obvious from the variations in period or from the observed timing residuals themselves. These periodic patterns continue back through the entire 13-year data set. The solid curves indicate the predictions of a fit of three harmonically-related sinusoids to $\dot{\Delta P}$; the fundamental period is $1,009 \pm 8$ days. In addition to modelling $\dot{\Delta P}$, these fitted parameters, as expected, yield a good representation of $\dot{\Delta P}$.

variations are caused by a magnetospheric interaction similar to that between Jupiter and Io (for example, see ref. 7), as this would require at least two of these planets to interact with the pulsar magnetosphere over distances comparable to the size of the Earth's orbit. Timing noise (for example, see ref. 8), believed to be due to random movements of the fluid within the pulsar, could not produce the sustained and regular pattern which we observe. A Tkachenko oscillation⁹ in the interior superfluid might have a period of the order of several hundred days, but provides no explanation for the variations in pulse shape.

The close relationship between the periodic changes in the beam shape and the rate of spin-down is simply explained by precession of the neutron star spin axis (for example, see ref. 10). This effect is rarely observed in celestial bodies, but will occur if the star is deformed so that its spin axis is not aligned with its angular momentum vector. This would cause cyclic changes, possibly with harmonics, in the inclination angle α between the spin and magnetic axes. The result will be periodic variations in pulse-profile morphology as the line-of-sight to the Earth passes through different cross-sections of the pulsar's radiation beam, and corresponding changes in the spin-down torque acting on the pulsar. Within the context of the rotating magnetic dipole model, the slow-down rate and the magnetic inclination angle are related by $\dot{P} = 8\pi^2 m^2 \sin^2 \alpha / (3c^3 I P)$, where c is the speed of light, I the moment of inertia and m the dipolar magnetic moment of the

neutron star. The observed 0.7% variation in $\dot{P}$ therefore implies a fractional change of similar magnitude in $\sin^2 \alpha$, or a variation in α of 0.3° for $\alpha \approx 60^\circ$. That is roughly one-tenth of the observed profile width of about 3° , and approximately the same magnitude as the observed change in width. That the shape changes are likely to be due to precession is supported by the fact that the only two other pulsars with well-measured long-term profile shape changes are the double-neutron-star binaries PSRs B1913+16 and B1534+12, which are undergoing precession due to general-relativistic spin-orbit coupling^{11,12}. PSR B1828-11 is somewhat different from these two pulsars in that it is solitary, and hence its precession cannot be influenced by another body. Although there have been suggestions¹³⁻¹⁵ that free precession might be the cause of possibly correlated timing behaviour and profile shape changes in other isolated radio pulsars, none but PSR B1828-11 shows strongly periodic variation in either pulse shape or rotation, let alone both. For instance, in PSR B1642-03 (ref. 13), the variation in shape is only marginally significant, and is not convincingly correlated with the aperiodic timing variations. Another family of neutron stars, the anomalous X-ray pulsars, are known to display 'wobles' in their slowdown rates (for example, ref. 16); perhaps free precession could account for these variations, making it unnecessary to invoke the magnetar-strength magnetic fields required by the 'radiative precession' theory¹⁷. Further observations of all classes of neutron stars may therefore help to shed light on the physics involved in the precession indicated by the behaviour of PSR B1828-11.

Theoretical studies of the neutron-star interior present a difficulty here. The rotation of the superfluid, which accounts for a large proportion of the moment of inertia of the star, is contained in an array of vortices. The occurrence of and recovery from the starquakes known as glitches have been successfully described by models in which vortices pinned to the stellar crust become unpinned during a glitch¹⁸. However, the vortex pinning, even if it is imperfect, will damp out free (untorqued) precession on timescales of several hundred precession periods (see refs 1 and 2). Nevertheless, in light of the clearly periodic behaviour of the timing parameters and profile shape of PSR B1828-11, a model will have to be devised which allows for some amount of vortex pinning as well as precession. Perhaps the key to understanding will lie in how the vortex array, whether partly or wholly pinned, responds to a variable external electromagnetic torque such as that which a precession would apply. In this case, the precession may be long-lived but perhaps not perfectly regular. □

Received 1 September 1999; accepted 22 May 2000.

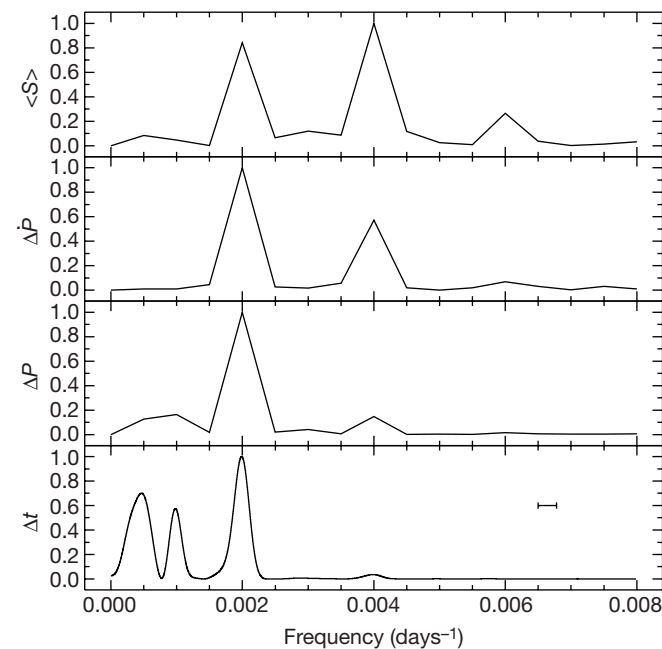


Figure 3 Harmonically related spectral features in the rotation and pulse shape of PSR B1828-11. The panels show the spectral power of the residuals Δt , ΔP , $\Delta \dot{P}$ and the mean pulse shape parameter $\langle S \rangle$. The normalizations are arbitrary. We used a one-dimensional CLEAN algorithm (for example, see ref. 19) for the timing residuals because of the uneven sampling within that data set; the horizontal bar indicates the full width of the smoothing function applied to the timing residual spectrum by this algorithm, and indicates that most of the spectral features are unresolved. The spectra exhibit harmonically related periodicities of approximately 1,000, 500 and 250 days, and confirm that the variations in shape and rotation are related. There is also a strong indication of the presence of a further harmonically related periodicity of approximately 167 days. The relative strengths of the harmonic components in the lower three plots change as expected from successive differentiation of the sinusoidal components of the timing residuals. We note that there is a good chance that the apparent power at around 1/2,000 days in the lowest plot could be due primarily to timing noise in this young pulsar; for this reason we do not include this periodicity in the discussion in the text.

- Shaham, J. Free precession of neutron stars: role of possible vortex pinning. *Astrophys. J.* **214**, 251–260 (1977).
- Sedrakian, A., Wasserman, I. & Cordes, J. M. Precession of isolated neutron stars. I. Effects of imperfect pinning. *Astrophys. J.* **524**, 341–460 (1999).
- Bailes, M., Lyne, A. G. & Shemar, S. L. in *Planets Around Pulsars* (eds Phillips, J. A., Thorsett, S. E. & Kulkarni, S. R.) 19–30 (Astronomical Society of the Pacific Conference Series, San Francisco, 1993).
- Wolszczan, A. & Frail, D. A. A planetary system around the millisecond pulsar PSR 1257+12. *Nature* **355**, 145–147 (1992).
- Rasio, F. A., Nicholson, P. D., Shapiro, S. L. & Teukolsky, S. A. An observational test for the existence of a planetary system orbiting PSR 1257+12. *Nature* **355**, 325–326 (1992).
- Wolszczan, A. Confirmation of Earth-mass planets orbiting the millisecond pulsar PSR 1257+12. *Science* **264**, 538–542 (1994).
- Kennel, C. F. & Coroniti, F. V. in *Solar System Plasma Physics*. Vol. 2 *Magnetospheres* 105–181 (North-Holland, Amsterdam, 1979).
- Lyne, A. G. in *Pulsars: Problems and Progress* (eds Johnston, S., Walker, M. A. & Bailes, M.) 73–81 (IAU Colloquium 160, Astronomical Society of the Pacific, San Francisco, 1996).
- Tkachenko, K. Stability of vortex lattices. *Sov. Phys. JETP* **23**, 1049–1056 (1966).
- Goldstein, H. *Classical Mechanics* 205–213 (Addison-Wesley, Reading, Massachusetts, 1980).
- Weisberg, J. M., Romani, R. W. & Taylor, J. H. Evidence for geodetic spin precession in the binary pulsar 1913+16. *Astrophys. J.* **347**, 1030–1033 (1989).
- Arzoumanian, Z. *Radio Observations of Binary Pulsars: Clues to Binary Evolution and Tests of General Relativity*. Thesis, Princeton Univ. (1995).
- Cordes, J. M. in *Planets around Pulsars* (eds Phillips, J. A., Thorsett, S. E. & Kulkarni, S. R.) 43–60 (Astronomical Society of the Pacific Conference Series Vol. 36, San Francisco, 1993).
- Suleymanova, S. A. & Shitov, Y. P. Secular variations of the integrated pulse profile of, and torque on, PSR 2217+47: Precession of the neutron star? *Astrophys. J. Lett.* **422**, 17–20 (1994).

15. D'Alessandro, F. & McCulloch, P. M. Long-term timing observations of four southern pulsars. *Mon. Not. R. Astron. Soc.* **292**, 879–886 (1997).
16. Mereghetti, S. A spin-down variation in the 6 second X-ray pulsar 1E 1048.1-5937. *Astrophys. J.* **455**, 598–602 (1995).
17. Melatos, A. Bumpy spin-down of anomalous X-ray pulsars: The link with magnetars. *Astrophys. J.* **519**, L77–L80 (1999).
18. Alpar, M. A., Anderson, P. W., Pines, D. & Shaham, J. Vortex creep and the internal temperature of neutron stars. I. General theory. *Astrophys. J.* **276**, 325–334 (1984).
19. Roberts, D. H., Lehar, J. & Dreher, J. W. Time series analysis with CLEAN. I. Derivation of a spectrum. *Astron. J.* **93**, 968–989 (1987).
20. Lyne, A. G. & Graham-Smith, F. *Pulsar Astronomy* 14–22 (Cambridge, Univ. Press, 1998).

Acknowledgements

We thank D. J. Nice for providing the two-standard-profile frequency-domain fitting routine, M. Goss for obtaining the interferometric pulsar position, F. Graham Smith for reading of the manuscript and F. Camilo for discussions. I.H.S. received support from a Natural Sciences and Engineering Research Council of Canada postdoctoral fellowship.

Correspondence and requests for materials should be addressed to A.G.L.
(e-mail: agl@jb.man.ac.uk).

Vortex-like excitations and the onset of superconducting phase fluctuation in underdoped $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$

Z. A. Xu^{*,†}, N. P. Ong^{*}, Y. Wang^{*}, T. Kakeshita[‡] & S. Uchida[‡]

^{*} Joseph Henry Laboratories of Physics, Princeton University, Princeton, New Jersey 08544, USA

[‡] School of Frontier Sciences, University of Tokyo, Yayoi 2-11-16, Bunkyo-ku, Tokyo 113-8656, Japan

Two general features of a superconductor, which appear at the critical temperature, are the formation of an energy gap and the expulsion of magnetic flux (the Meissner effect). In underdoped copper oxides, there is strong evidence that an energy gap (the pseudogap¹) opens up at a temperature significantly higher than the critical temperature (by 100–220 K). Certain features of the pseudogap suggest that it is closely related to the gap that appears at the critical temperature (for example, the variation of the gap magnitudes around the Fermi surface and their maximum amplitudes are very similar^{2,3}). However, the Meissner effect is absent in the pseudogap state. The nature of the pseudogap state, and its relation (if any) to the superconducting state are central issues in understanding copper oxide superconductivity. Recent evidence suggests that, in the underdoped regime, the Meissner state is destroyed above the critical temperature by strong phase fluctuations^{4–7} (as opposed to a vanishing of the superfluid density). Here we report evidence for vortices (or vortex-like excitations) in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ at temperatures significantly above the critical temperature. A thermal gradient is applied to the sample in a magnetic field. Vortices are detected by the large transverse electric field produced as they diffuse down the gradient (the Nernst effect). We find that the Nernst signal is anomalously enhanced at temperatures as high as 150 K.

In conventional superconductors, fluctuations in the phase $\theta(r)$ of the superconducting wavefunction $\psi \exp[i\theta(r)]$ incur a sizeable cost in energy (the phase stiffness energy is large). Hence $\theta(r)$ is uniform in the absence of field and currents. In the underdoped

copper oxides, however, the small superfluid density n_s implies a small phase stiffness energy⁴. The Meissner state is readily destroyed by strong phase fluctuations. This suggests that rapidly diffusing vortices, which are highly effective in destroying phase coherence, may help to limit the critical temperature T_c . In a recent experiment, Corson *et al.*⁷ measured the complex conductivity in underdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ (BSCCO) at terahertz frequencies, and deduced a Kosterlitz–Thouless transition⁸ at T_c , with a fluctuating regime that extends about 25 K above T_c .

To detect directly the presence of vortices above T_c , we have adopted a different approach. By using a strictly d.c. probe that selectively senses vortex motion, we may detect excitations that are long-lived. An applied thermal gradient ($-\nabla T$) causes vortices to diffuse with a velocity $\mathbf{v}$ ($\mathbf{v}$ may be tilted relative to $-\nabla T$). The distinguishing feature of moving vortices is the Josephson electric field given by $\mathbf{E} = \mathbf{B} \times \mathbf{v}$, which reflects phase slippage caused by moving vortices. The y -component of the $\mathbf{E}$ -field leads to a very large Nernst effect. With $-\nabla T \parallel \mathbf{x}$, the Nernst coefficient is defined as

$$\nu = E_y / (B |\nabla T|) \quad (1)$$

with the induction field $\mathbf{B} \parallel \mathbf{z}$. (Previous Nernst experiments in the copper oxides were restricted to optimally doped samples^{9–11}.) In our experiment, we measured ν versus T in five crystals of $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ (LSCO; samples 1–5), with $0.05 < x < 0.17$, and in a crystal of Nd-doped $\text{La}_{2-y}\text{Nd}_y\text{Sr}_x\text{CuO}_4$ (sample N).

Figure 1 displays traces of the Nernst signal E_y versus the applied field H in sample 3 ($x = 0.10$) at temperatures between 12 and 35 K (Fig. 1a) and above 35 K (Fig. 1b). At 12 and 15 K, the behaviour

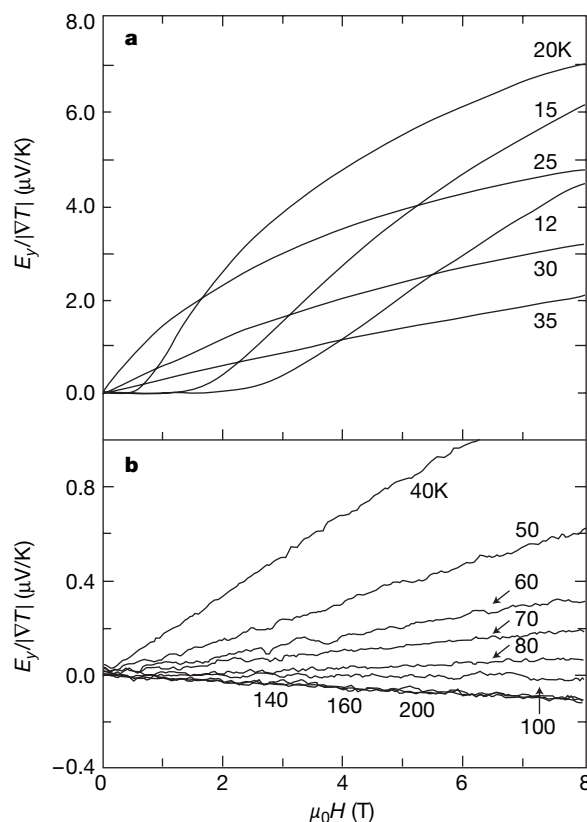


Figure 1 Nernst signals. **a**, The Nernst signal E_y (normalized to unit gradient) versus $H \parallel \mathbf{c}$ in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ (sample 3, $x = 0.10$) at temperatures 12–35 K. **b**, The Nernst signal from 40 to 200 K. Above 20 K, the applied gradient is 5 K cm^{-1} , while below 20 K, it is half as large. When vortex pinning is large ($T < 25 \text{ K}$), E_y is zero over a range of $H < H_m$. Above 140 K, the curves tend asymptotically to a straight line of negative slope.

[†] Present address: Department of Physics, Zhejiang University, Hangzhou 310027, China.

of E_y versus H resembles the flux-flow resistivity. When H exceeds a T -dependent field scale H_m , the onset of vortex motion produces a large Nernst signal that increases rapidly with H in accordance with $\mathbf{E} = \mathbf{B} \times \mathbf{v}$. The surprising feature is that the Nernst signal is a sizeable fraction of the low- T values above $T_c = 28$ K (see curves at 30 and 35 K). At higher T (Fig. 1b), the signal decreases, gradually approaching a straight line of negative slope, which we identify with the background signal ν_n from the holes. Similar traces are obtained in all samples studied. The Nernst coefficient ν , calculated from the initial slope of E_y versus H , is shown for samples 1–5 in log–log scale in Fig. 2 (with the background ν_n subtracted). In each sample, the maximum value of ν ($\sim 2 \mu\text{V}/\text{KT}$) is similar to that in $\text{YBa}_2\text{Cu}_3\text{O}_7$ (YBCO)¹⁰. With increasing T , ν falls by two decades from its peak value to the normal-state value ν_n . Comparing across samples, we noticed a systematic trend of the rate of decrease of ν above T_c (arrows). As x decreases from 0.17 to 0.05, ν falls less and less steeply. To display this trend more clearly, we show in expanded (linear) scale the behaviour of ν at high T (Fig. 3a). The arrows indicate the onset temperature T_{onset} at which we resolve the anomalous signal from the background ν_n (broken horizontal lines). As x decreases from 0.17 to 0.05, the interval between T_c and T_{onset} in each sample systematically expands. In the most underdoped samples (1 and 2), the Nernst signal is anomalously enhanced up to about 150 K. The enhanced signal presents intriguing evidence for vortices or vortex-like excitations that persist to elevated temperatures.

For comparison, we discuss the ‘weak’ Nernst background originating from the action of the Lorentz force on the diffusing holes (that is, in the absence of vortices). In an $\mathbf{E}$ -field and a gradient $-\nabla T$, the current density is comprised of two terms: $\mathbf{J} = \alpha(-\nabla T) + \sigma \cdot \mathbf{E}$, where σ is the two-dimensional electrical conductivity tensor, and α the two-dimensional conductivity tensor. In a magnetic field $\mathbf{B} \parallel \mathbf{z}$, the two terms produce Hall currents

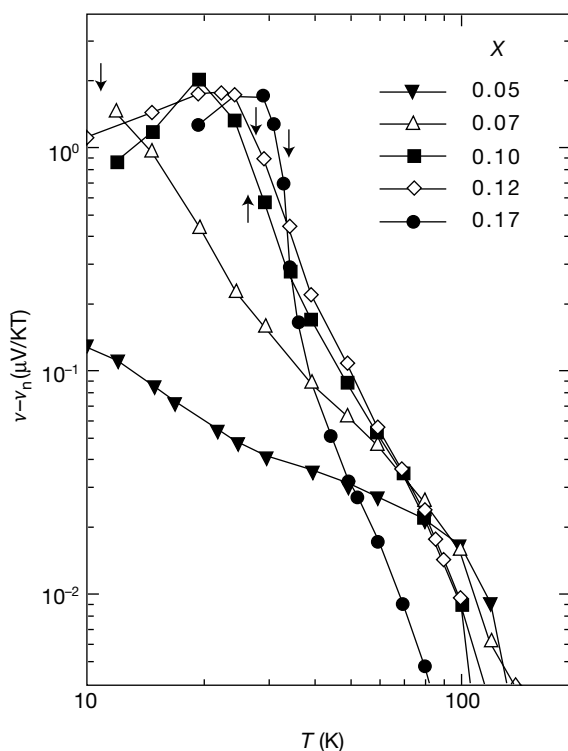


Figure 2 Log–log plot of the anomalous part of the Nernst coefficient. $\nu - \nu_n$ versus T in LSCO is plotted for samples 1–5, with $x(T_c) = 0.05$ (<4 K), 0.07 (11 K), 0.10 (27.5 K), 0.12 (29 K) and 0.17 (35.4 K), respectively. As x decreases, the rate of decrease of ν above T_c (arrows) decreases, while the onset temperature increases.

along $\mathbf{y}$ of opposite signs which result in a strongly reduced Nernst coefficient¹² (see legend of Fig. 3):

$$E_y/(B|\nabla T|) = [(\alpha_{xy}/\sigma) - (\alpha/\sigma)(\sigma_{xy}/\sigma)]/B \quad (2)$$

where σ and α are the diagonal elements of σ and α , respectively. The second term may be re-cast as $-S \tan\theta_H/B$, with S the thermopower and $\tan\theta_H$ the Hall angle. We verify that the cancellation between the two terms is nearly complete in LSCO. In each sample, we have measured S and σ_{xy} separately in order to calculate $S \tan\theta_H/B$. With decreasing T , $S \tan\theta_H/B$ displays a broad maximum (solid symbols in Fig. 3a). In comparing its value with ν (in sample 2 above 130 K, for example), it is clear that the latter is about 30 times

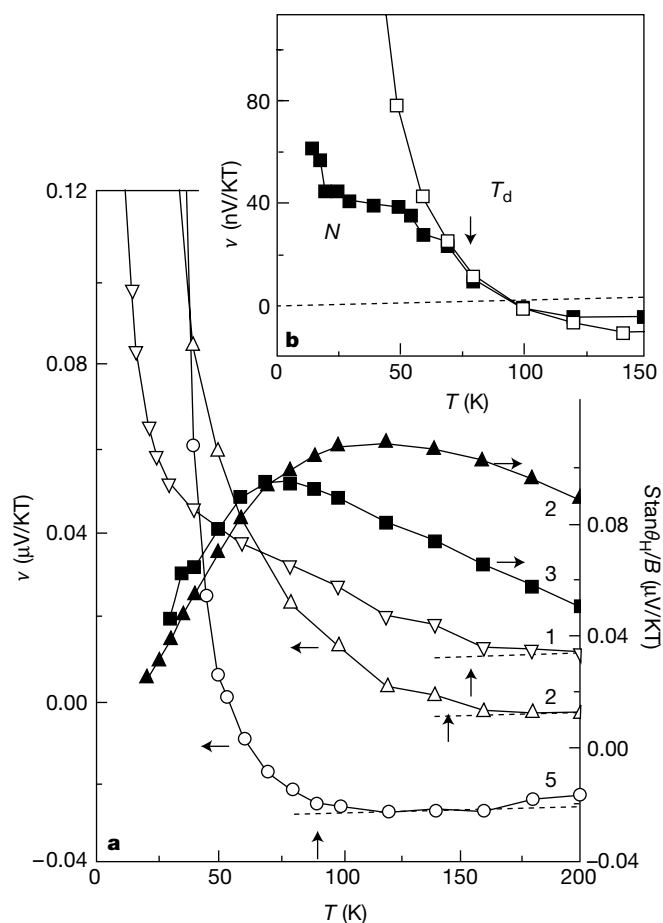


Figure 3 Nernst coefficients versus T . **a**, Expanded view of ν versus T in samples 1 (down triangles), 2 (open up triangles) and 5 (circles), showing the steep divergence of the vortex signal from the background Nernst signal (shown as broken lines). **b**, Comparison of ν in sample 3 ($x = 0.10$, open symbols) with ν in Nd-doped LSCO (sample N, with $y = 0.60$ and $x = 0.10$, solid symbols). In **a**, arrows indicate T_{onset} . The solid symbols represent the contribution $S \tan\theta_H/B$ to ν in samples 2 and 3 (see equation (2)) obtained by measuring S and $\tan\theta_H$ independently. Above T_{onset} , cancellation between the two terms in equation (2) suppresses ν to values 10–30 times smaller than $S \tan\theta_H/B$. In the absence of vortices, the experimental constraint $J_x = 0$ leads to two equal and opposite currents parallel to $\mathbf{x}$: that is, $\alpha(-\partial_x T) + \sigma E_x = 0$ (whence $S = \alpha/\sigma$). In finite $\mathbf{H} \parallel \mathbf{z}$, the fields $(-\partial_x T)$ and E_x generate transverse Hall currents (along $\mathbf{y}$) of opposite signs. The constraint $J_y = 0$ leads to $\alpha_{yx}(-\partial_x T) + \sigma_{yx} E_x + \sigma E_y = 0$ (we neglect $-\partial_y T$, which is very small in copper oxides). Near-cancellation between the first two terms leads to a small E_y , which is observed as the Nernst signal (equation (2)). The data show that the third term is 10–30 times smaller than either of the other two. The comparison in **b** highlights the sensitivity of the divergent Nernst signal to the structural transition occurring at T_d . ν in samples 3 and N are nominally equal from 150 K to $T_d = 74$ K (arrow). Below T_d , ν remains near 40 nV/KT in sample N, whereas it increases by a further factor of 50 in sample 3.

smaller, that is, the two terms in equation (2) differ by only 1 part in 30. Hence, the background Nernst signal is weakly T dependent, and about 100–500 times weaker than the peak vortex signal. The comparison shows that the possibility is very remote that the divergent Nernst signal arises from the charge carriers.

As a further test, we compare the Nernst signal in sample 3 ($x = 0.10$) with that in a Nd-doped LSCO crystal (sample N) with the same x (see Fig. 3 inset). The Nd impurities induce a structural transition at $T_d = 74$ K. Charge ordering below T_d , consistent with static stripe formation¹³, strongly suppresses the onset of the Meissner state to below 5 K. Hence, we expect superconducting fluctuations to be strongly suppressed below T_d . If the divergent component of the Nernst signal in Nd-free samples indeed originates from vortices, it should also be markedly affected below T_d . Our measurements (Fig. 3 inset) show that, from 150 K to T_d , ν in the Nd-doped crystal increases rapidly (and is very close to the value in sample 3). Below T_d , however, ν in sample N remains at the value of 40 nV/KT, whereas it grows by a further factor of 50 in sample 3. The comparison confirms that when superconducting fluctuations are strongly suppressed below T_d , there is a near-complete suppression of the divergent part of ν . This provides strong support for our identification of the divergent part of ν with a vortex-like excitation clearly related to superconducting pairing, but occurring up to 120 K above the Meissner onset at T_c .

Although we believe that the evidence for such identification is very strong, it remains to be confirmed by more direct means, such as imaging. Moreover, the finding raises the following two questions: (1) Is it confined to the LSCO family? We have nearly completed similar measurements in underdoped $\text{YBa}_2\text{Cu}_3\text{O}_y$ and Zn-doped $\text{YBa}_2(\text{Cu}_{3-z}\text{Zn}_z)\text{O}_y$ (ref. 14). Underdoped YBCO ($y = 6.40$ and 6.60) and Zn-doped YBCO show enhanced Nernst signals extending to 100–130 K, broadly similar to those reported here. The exception is over-doped YBCO ($y = 7.0$), in which the Nernst signal vanishes abruptly above about 95 K. (2) Are there other explanations for the enhanced Nernst signal? For convenience,

we divide the data into a low- T region in which ν strongly diverges towards T_c , and a high- T region where it smoothly tends asymptotically to ν_n as T increases (Fig. 3). As we argued from the data in Fig. 3b, the low- T divergence (below T_d) is suppressed in parallel with superconductivity, so it must be intimately related to Cooper pair formation. Above T_d (74 K), however, this test cannot be invoked to exclude a non-vortex origin; for example, the cancellation in equation (2) may be spoiled by dynamic stripe formation. This objection is best addressed experimentally, by doping with Zn (in YBCO, Zn is highly effective (Y. W. *et al.*, unpublished work) in removing all the anomalous signal at high T).

Finally, we cannot exclude the interesting possibility that the enhanced signal originates from an unsuspected novel excitation in the pseudogap state that binds a fluxoid, and responds to the thermal gradient. This question merits further study.

With these caveats in mind, we display as contour plots the anomalous signal ($\nu - \nu_n$) in the phase diagram of LSCO (broken lines in Fig. 4). The plot provides an overview of how high in T the anomalous signal extends for various x . The smaller x is, the higher the signal extends, as noted above. However, the large-amplitude contours (for 1 $\mu\text{V/KT}$, say) tend to follow the T_c line. We define the contour set by our resolution (4 nV/KT) as T_{onset} . These features strongly suggest that, although T_{onset} measures the highest temperature of the vortex-like excitations, the Meissner state is stable only when ν exceeds 1–2 $\mu\text{V/KT}$ (this is also evident in Fig. 2). Values of the pseudogap T^* in LSCO are less certain than in YBCO. From heat capacity measurements^{1,15}, $T^* \approx 350$ K at $x = 0.08$, and falls linearly to ~ 90 K at $x = 0.22$. With this estimate, T_{onset} is only half as large as T^* . Over the doping range investigated, T_{onset} decreases with x as the pseudogap temperature. This suggests that T_{onset} and T^* may be controlled by the same energy scale. $\square$

Received 6 March; accepted 27 May 2000.

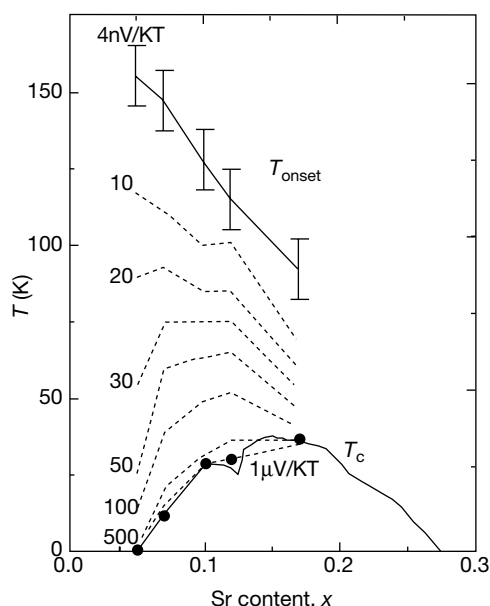


Figure 4 Contour plot of $(\nu - \nu_n)$ versus x in the phase diagram of LSCO. The contour plot displays how high in T the vortex-like excitations extend for each value of x . The upper solid line T_{onset} is the contour set by our resolution. The pseudogap T^* estimated from heat capacity¹⁵ is about a factor of two larger than T_{onset} . Values of T_c in our samples (circles) match the T_c line (lower solid line) from Takagi *et al.*¹⁴ We note that the T_c line is roughly similar to the contour line $\nu = 1 \mu\text{V/KT}$.

- Timusk, T. & Statt, B. The pseudogap in high-temperature superconductors: an experimental survey. *Rep. Prog. Phys.* **62**, 61–122 (1999).
- Loeser, A. G. *et al.* Excitation gap in the normal state of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Science* **273**, 325–329 (1996).
- Ding, H. *et al.* Spectroscopic evidence for a pseudogap in the normal state of underdoped high- T_c superconductors. *Nature* **382**, 51–54 (1996).
- Emery, V. J. & Kivelson, S. A. Importance of phase fluctuations in superconductors with small superfluid density. *Nature* **374**, 434–437 (1995).
- Norman, M., Randeria, M., Ding, H. & Campuzano, J. C. Phenomenology of the low-energy spectral function in high- T_c superconductors. *Phys. Rev. B* **57**, 11093–11096 (1998).
- Lee, P. A. & Wen, X. G. Unusual superconducting state of underdoped cuprates. *Phys. Rev. Lett.* **78**, 4111–4114 (1997).
- Corson, J., Mallozzi, R., Orenstein, J., Eckstein, J. N. & Bozovic, I. Vanishing of phase coherence in underdoped $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Nature* **398**, 221–223 (1999).
- Ambegaokar, V., Halperin, B. I., Nelson, D. R. & Siggia, E. D. Dynamics of superfluid films. *Phys. Rev. B* **20**, 1806–1826 (1980).
- Hagen, S. J., Lobb, C. J., Greene, R. L., Forrester, M. G. & Talvacchio, J. Flux-flow Nernst effect in epitaxial $\text{YBa}_2\text{Cu}_3\text{O}_7$. *Phys. Rev. B* **42**, 6777–6780 (1990).
- Ri, H. C. *et al.* Nernst, Seebeck, and Hall effect in the mixed state of $\text{YBa}_2\text{Cu}_3\text{O}_{7+\delta}$ and $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ —a comparative study. *Phys. Rev. B* **50**, 3312–3329 (1994).
- Clayhold, J. A., Linnen, A. W. Jr, Chen, F. & Chu, C. W. Normal-state Nernst effect in a $\text{Tl}_2\text{Ba}_2\text{CaCu}_2\text{O}_{8+\delta}$. *Phys. Rev. B* **50**, 4252–4255 (1994).
- Sondheimer, E. H. The theory of the galvanomagnetic and thermomagnetic effects in metals. *Proc. R. Soc. Lond. A* **193**, 484–512 (1948).
- Tranquada, J. M., Sternlieb, B. J., Axe, J. D., Nakamura, Y. & Uchida, S. Evidence for stripe correlations of spins and holes in copper oxide superconductors. *Nature* **375**, 561–563 (1995).
- Takagi, H., Ido, T., Ishibashi, S., Uota, M. & Uchida, S. Superconductor-to-nonsuperconductor transition in $(\text{La}_{1-x}\text{Sr}_x)_2\text{CuO}_4$ as investigated by transport and magnetic measurements. *Phys. Rev. B* **40**, 2254–2261 (1989).
- Loram, J. W., Mirza, K. A., Cooper, J. R., Athanassopoulou, N. & Liang, W. Y. in *Proc. 10th HTS Anniversary Workshop on Physics, Materials and Applications* (eds Batlogg, B., Chu, C. W., Chu, W. K., Gubser, D. U. & Muller, K. A.) 341 (World Scientific, Singapore, 1996).

Acknowledgements

Comments by P. W. Anderson, P. A. Lee and Z. Weng are acknowledged. The research is supported by the US Office of Naval Research and the US National Science Foundation. N.P.O. and S.U. acknowledge support from the New Energy and Industrial Technology Development Organization, Japan.

Correspondence and requests for materials should be addressed to N.P.O. (e-mail: npo@princeton.edu).

Pixel switching of epitaxial Pd/YH_x/CaF₂ switchable mirrors

J. W. J. Kerssemakers, S. J. van der Molen, N. J. Koeman, R. Günther & R. Griessen

Division of Physics and Astronomy, Faculty of Sciences, Vrije Universiteit, De Boelelaan 1081, NL-1081 HV Amsterdam, The Netherlands

Exposure of rare-earth films to hydrogen can induce a metal–insulator transition, accompanied by pronounced optical changes. This ‘switchable mirror’ effect¹ has received considerable attention from theoretical^{2–4}, experimental⁵ and technological⁶ points of view. Most systems use polycrystalline films, but the synthesis of yttrium-based epitaxial switchable mirrors⁷ has also been reported. The latter form an extended self-organized ridge network during initial hydrogen loading⁷, which results in the creation of micrometre-sized triangular domains. Here we observe homogeneous and essentially independent optical switching of individual domains in epitaxial switchable mirrors during hydrogen absorption. The optical switching is accompanied by topographical changes as the domains sequentially expand and contract; the ridges block lateral hydrogen diffusion and serve as a microscopic lubricant for the domain oscillations. We observe the correlated changes in topology and optical properties using *in situ* atomic force and optical microscopy. Single-domain phase switching is not observed in polycrystalline films, which are optically homogeneous⁸. The ability to generate a tunable, dense pattern of switchable pixels is of technological relevance for solid-state displays based on switchable mirrors.

In 1996, Huiberts *et al.* discovered that the optical and electronic properties of thin polycrystalline films of Y and La change dramatically when exposed to hydrogen gas¹. For example, YH_x changes from a reflective h.c.p.-metal (α -YH_x, $x < 0.21$) to a very weakly transparent f.c.c.-metal (β -YH₂ ± ϵ) and finally, to a transparent h.c.p.-semiconductor (γ -YH₃ ± δ) upon hydrogenation. The β – γ

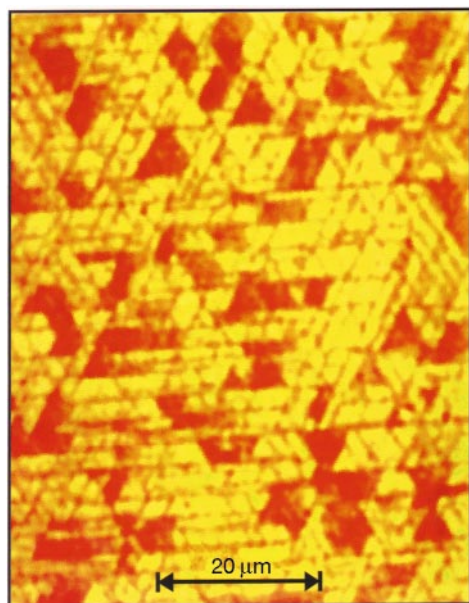


Figure 1 Optical domain switching. Real-colour optical micrograph in transmitted light of a 400-nm-thick Y film capped with 7 nm of Pd. After loading to the trihydride phase YH₃ ± δ , the hydrogen content is slowly lowered. The yellowish transparent YH₃ ± δ film switches back domain by domain to reddish opaque YH₂. The domains are bounded by a triangular network of reddish opaque lines.

transition is reversible at room temperature. The hydrides of all trivalent rare-earth metals exhibit similar behaviour⁹. Several theoretical studies have modelled the metal–insulator transition^{2,3,4}, but consensus on its exact nature has not yet been reached¹⁰. Some propose a Peierls distortion mechanism² while others invoke strong electron-correlation effects and introduce a new, ‘breathing’ Hubbard model^{3,4}.

We prepare epitaxial switchable mirrors by electron-gun deposition of yttrium on (111) oriented CaF₂ substrates⁷ at elevated temperatures (300–700 °C) in ultra-high vacuum (UHV, base pressure 10^{−9} mbar). To promote hydrogen dissociation and to protect the highly reactive Y against oxidation, the films are capped with a Pd layer thin enough (1–30 nm) to be optically transparent. The advantage of CaF₂ is that, unlike other substrates which are metallic¹¹ or require a metallic buffer layer¹², it is transparent and insulating and is therefore ideally suited for investigation of electrical and optical properties of epitaxial switchable mirrors. In these films we found a highly ordered out-of plane stacking of the hexagonal (111) planes⁷. This stacking remains unaffected upon hydrogenation^{7,12}. However, after the first loading from Y to YH₂ a dense ridge network reflecting the sixfold symmetry of the lattice is created^{7,13}. With the lattice orientation as determined by X-ray and Rutherford backscattering spectroscopy (RBS) channelling, we conclude the ridges to be along the {1120} directions. The protruding volume of the ridges matches the excess material

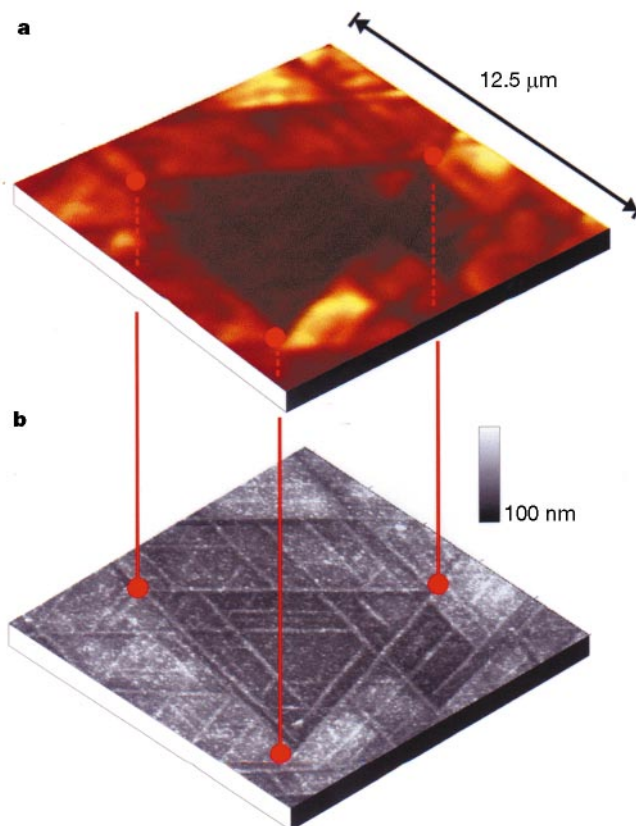


Figure 2 One-to-one correlation between optical and structural texture. **a**, Optical micrograph in transmitted light of a dihydride/trihydride mixed-phase region of a 150-nm-thick Y/CaF₂ film capped with 20 nm of Pd. The image is 12.5 × 12.5 μm². The large triangle is almost entirely in the electrically conducting and opaque dihydride phase (YH₂). The bright yellow regions are in the transparent trihydride phase (YH₃). **b**, AFM micrograph of the same area as in **a**. A triangular network of ridges bounds micrometre-sized flat domains of various heights (dark is low, white is high). The surface topography closely matches the optical pattern in **a**, indicating that domains are in different stages of switching from YH₂ (opaque, contracted domains) to YH₃ (transparent, expanded domains).

squeezed out by the expected in-plane expansion of 0.8% of the film when loaded for the first time from Y to YH_2 . We note that this triangular ridge network does not show any further changes upon cyclic (un)loading between YH_2 and YH_3 .

Hydrogen loading and unloading of epitaxial yttrium films is studied *in situ* with an Olympus BX60F5 optical microscope and a Nanoscope III Tapping Mode/Contact Mode atomic force microscope (AFM). Both are equipped with a hydrogen-gas loading chamber. Additional *ex situ* measurements are done with Nomarski imaging in the Olympus microscope.

A typical example of the regular triangular optical pattern which occurs in epitaxial switchable mirrors is given in Fig. 1 for a 400-nm-thick Y film capped with 20 nm of Pd. The film, originally loaded to $\text{YH}_{3-\delta}$ at 1 bar H_2 , is imaged in transmitted light while it is slowly dehydrogenating. A dense, micrometre-spaced network of opaque lines oriented along the main, hexagonal directions of the epitaxial Y film is observed. The lines define domains that are either yellowish transparent, or reddish opaque. It suggests a spectacular discretization of mirror switching in separate domains, an effect that is not at all observed in polycrystalline switchable mirrors. A remarkable characteristic of domain switching is shown in Fig. 2, where a direct comparison is made between optical transmission (Fig. 2a) and topographical texture (Fig. 2b) of one specific area of a 150-nm-thick YH_x film. The sharply bounded domains that differ strongly in transparency are also found in the same area using AFM. There is a surprisingly clear correspondence between the height of the domains and their optical transmission. This can be understood considering that the transition from β -dihydride to γ -trihydride

does not only involve an optical change from reddish opaque to yellowish transparent^{8,14}, but also an out-of-plane axis expansion of 9.3%, in good agreement with the observed topography change^{7,12,15}. We therefore conclude that β - γ -switching on an epitaxial yttrium mirror proceeds in a domain-wise fashion, creating a 'Manhattan skyline' of expanded and contracted skyscrapers.

The ridges at the domain boundaries are optically visible as dark lines (see also Fig. 1), indicating a switching behaviour different from that of the enclosed areas. This is confirmed by AFM measurements of a domain during switching, as shown in Fig. 3. Figure 3a is an AFM scan of a partially unloaded YH_x film ($x \approx 2.6$). Again, the triangular network is visible, now as a set of protruding, connected ridges, forming domains in between. We focus on the centre domain, which is clearly lowered with respect to its surroundings. During reloading of the film in a H_2 atmosphere an AFM line scan is continuously recorded as a function of time (Fig. 3c). The originally lower domain expands as a whole shortly after hydrogen is absorbed, until the level of the neighbouring domains is reached (Fig. 3b) in approximately 20 s. During this, the ridge network stays essentially unchanged. The relative height shift (here about 20 nm) between neighbouring domains implies that the ridges function as microscopic lubrication borders. Such ridge-confined lubrication explains the fact that on average, an epitaxial film remains undamaged during cyclic hydrogenation between the β -($\text{YH}_{1.8-2.1}$) and γ -($\text{YH}_{3-\delta}$) phases, despite the enormous local shear. The simultaneous occurrence of optical and topographical switching is at the heart of the pixel-wise Manhattan switching.

One should realize that Manhattan switching is not encountered in usual two-phase systems, which involve mobile phase boundaries, while on epitaxial switchable mirrors, the boundaries between β - and γ -phase domains are strongly locked to the triangular ridges. This implies that ridges act as barriers for lateral hydrogen diffusion. We verified this in lateral diffusion experiments (see refs 5 and 16). The observation of Manhattan switching is only possible because of the intrinsic large size of the triangular domains in epitaxial films. However, all the average properties of epitaxial films are identical to

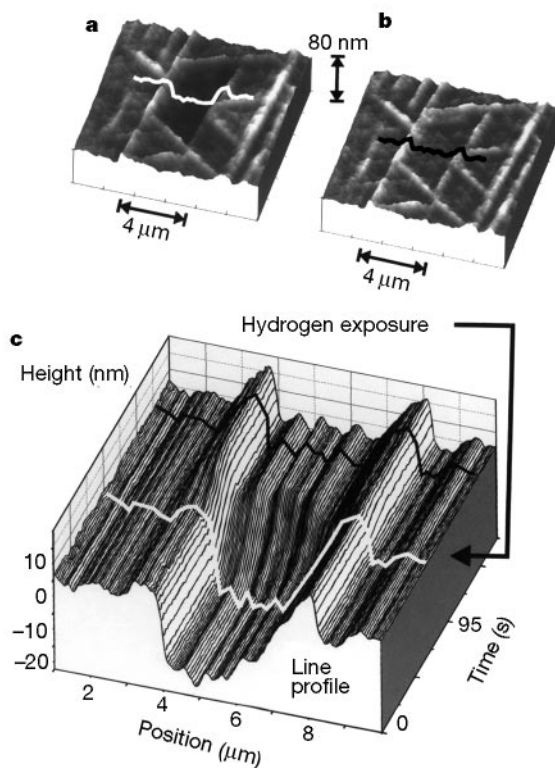


Figure 3 Structural pixel switching of a domain. **a**, AFM micrograph of a partially unloaded $\text{YH}_{3-\delta}$ film, showing domains bounded by a triangular array of ridges. One domain is clearly lowered, owing to hydrogen loss that is not replenished from neighbouring trihydride domains. **b**, After *in situ* re-exposure of the film to hydrogen, the central domain reloads to YH_3 , resulting in a height increase. **c**, A line scan taken at the white and black solid lines in **a** and **b** reflects the time evolution of the structural switching. We note that, although the height of the central domain increases by about 20 nm, the surrounding domains are totally unaffected. This implies that ridges act simultaneously as hydrogen barriers and lubrication walls.

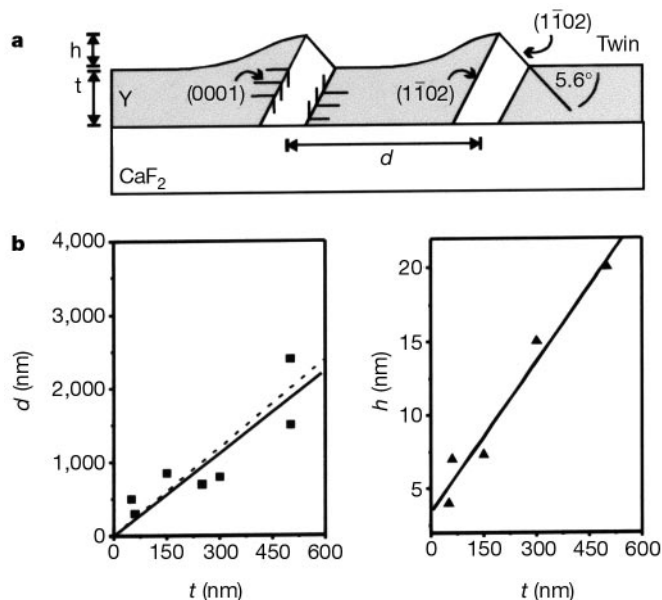


Figure 4 Tunability of the domain size. **a**, Schematic representation (vertical scale is exaggerated) of the ridge pattern characterized by a typical spacing d and ridge height h and presumed to be twinning defects caused by in-plane expansion. The hexagonal (0001) planes are marked as solid lines. **b**, Both d and h depend linearly on film thickness t (the solid lines are fits). The thickness dependence of the ridge spacing can be understood in terms of a model involving a partially clamped film, in which the stress relaxation caused by one ridge has a limited functional range (dashed line, see main text).

those of polycrystalline switchable mirrors evaporated on sapphire⁸ (A. T. M. van Gogh, manuscript in preparation). Such films are not epitaxial but only show a texture with the *c*-axis out of plane. Although *in situ* X-ray studies revealed the existence of two coexisting phase regions ($\text{Y}-\text{YH}_2$ and YH_2-YH_3), no optical inhomogeneity was observed⁸. This is most likely to be due to the small size of their domain structure^{17,18}.

The self-organized, switchable tile pattern of epitaxial switchable mirrors suggests potential applications. A switchable mirror can be integrated in an all-solid-state device⁶ (P. van der Sluis *et al.*, manuscript in preparation), so the switching tiles could be treated as individually addressable pixels in an optical display. Furthermore, the reversible, dramatic out-of-plane height variation (9.3%) offers potential for high-load (in the order of GPa) micromechanical actuators of far smaller dimensions than devices based on piezoelectric expansion.

One of the prerequisites for such devices is a dense and regular pixel pattern with a tunable domain size. One way to tune the domain size follows from Fig. 4 which shows that ridge spacing *d* and ridge height *h* increase linearly with film thickness *t*. The observed trend and magnitude can be understood in terms of a partially clamped film, in which the stress relaxation caused by one ridge has a limited functional range. The ridge network is created by the 0.8% in-plane expansion during the $\text{Y}-\text{YH}_2$ transition: the constraint by the substrate leads to large compressive stresses which are released by pushing some material outward in the form of ridges⁷. The measured volume of the ridges accounts for all the excess material. At the same time, the relatively wide spacing (in the order of μm) of the ridges implies significant in-plane material transport. This occurs through gliding of the plastic metallic film over the rigid substrate and release of the excess material at the ridges by plastic deformation. The onset of film sliding is determined by a critical shear γ_i between the film and the substrate. Analogously, the onset of ridge creation is determined by a critical stress σ_r . As soon as a given ridge is formed, it instantaneously lowers in its vicinity the average compressive stress $\sigma(x)$, just as a crack reduces local tensile stress¹⁹. Assuming a constant critical interface shear γ_i , the stress $\sigma(x)$ increases linearly away from the ridge, until it reaches the critical value σ_r at a certain distance. If the stress relief at the ridge is maximum, it follows that one ridge has a limited functional range around it, given by $d = 2t(\sigma_r/\gamma_i)$: the separation between the ridges *d* depends on the film thickness *t* scaled by the ratio $2\sigma_r/\gamma_i$. For any chosen crystal plane at an angle α with the film plane, σ_r corresponds to a critical shear $\gamma_p = \sigma_r \cos(\alpha) \sin(\alpha)$ along this plane. This yields $d = Ct$ with $C = 4(\gamma_p/\gamma_i)/\sin(2\alpha)$. Assuming $\gamma_p \approx \gamma_i$, we find that $C \geq 4$ for any plane. For hexagonal yttrium, a plausible choice is the (1102) twinning plane, at an angle α of 42.2° with the film surface (using measured lattice spacings⁷). This particular plane yields $C = 4$, consistent with the measured value of 3.7 ± 0.6 (see Fig. 4b). Moreover, a double twin (see Fig. 4a) creates asymmetric ridges with a steepest slope of $2(45 - \alpha) = 5.6$ degrees, in nice agreement with the measured (see also Fig. 3) average steepest slope of (5.8 ± 0.2) degrees.

Thus a fundamentally and technologically very attractive behaviour occurs in epitaxial switchable mirrors: single-domain phase switching. The effect takes place because the first loading of an epitaxial YH_2/CaF_2 film generates a self-organized, closed ridge structure which blocks hydrogen diffusion and accommodates completely plastic deformation. The domains defined by the ridge network switch homogeneously and independently. The possibility of generating a tunable, dense pattern of switchable pixels opens the way to solid state displays based on switchable mirrors. □

Received 7 February; accepted 8 June 2000.

- Huiberts, J. N. *et al.* Yttrium and lanthanum hydride films with switchable optical properties. *Nature* **380**, 231–234 (1996).
- Kelly, P. J., Dekker, J. P. & Stumpf, R. Theoretical prediction of the structure of insulating YH_3 . *Phys. Rev. Lett.* **78**, 1315–1318 (1997).

- Ng, K. K., Zhang, F. C., Anisimov, V. I. & Rice, T. M. Theory for metal hydrides with switchable optical properties. *Phys. Rev. B* **59**, 5398–5413 (1999).
- Eder, R., Pen, H. A. & Sawatzky, G. A. Kondo-lattice-like effects of hydrogen in transition metals. *Phys. Rev. B* **56**, 10115–10120 (1997).
- den Broeder, F. J. A. *et al.* Visualization of hydrogen migration in solids using switchable mirrors. *Nature* **394**, 656–658 (1998).
- Armitage, R., Rubin, M., Richardson, T., O'Brien, N. & Chen, Y. Solid-state gadolinium–magnesium hydride optical switch. *Appl. Phys. Lett.* **75**, 1863–1865 (1999).
- Nagengast, D. G., Kerssemakers, J. W. J., van Gogh, A. T. M., Dam, B. & Griessen, R. Epitaxial switchable yttrium-hydride mirrors. *Appl. Phys. Lett.* **75**, 1724–1726 (1999).
- Kremers, M. *et al.* Optical transmission spectroscopy of switchable yttrium hydride films. *Phys. Rev. B* **57**, 4943–4949 (1998).
- van der Sluis, P., Ouwerkerk, M. & Duine, P. A. Optical switches based on magnesium lanthanide alloy hydrides. *Appl. Phys. Lett.* **70**, 3356–3358 (1997).
- van Gogh, A. T. M., Kooij, E. S. & Griessen, R. Isotope effects in switchable metal-hydride mirrors. *Phys. Rev. Lett.* **83**, 4614–4617 (1999).
- Hayoz, J. *et al.* Observation of the reversible H-induced structural transition in thin Y films via x-ray photoelectron diffraction. *Phys. Rev. B* **58**, R4270–R4273 (1998).
- Remhof, A. *et al.* Hydrogen and deuterium in epitaxial Y(0001) films: Structural properties and isotope exchange. *Phys. Rev. B* **59**, 6689–6699 (1999).
- Grier, E. J. *et al.* Structural changes to epitaxial (0001) holmium layers during hydrogen loading. *J. Phys. D* **33**, 894–900 (2000).
- Kooij, E. S., van Gogh, A. T. M. & Griessen, R. In situ resistivity measurements and optical transmission and reflection spectroscopy of electrochemically loaded switchable YH_3 films. *J. Electrochem. Soc.* **146**, 2990–2994 (1999).
- Vajda, P. in *Handbook on the Physics and Chemistry of Rare Earths* Ch. 20 (eds Gschneider, K. A. & Eyring, L.) 207–291 (Elsevier, Amsterdam, 1995).
- Remhof, A. *et al.* In-situ x-ray diffraction topography studies on the phase formation in thin yttrium hydride films. *Phys. Rev. B* (in the press).
- van der Molen, S. J. *et al.* Hydriding kinetics of Pd capped YH_x switchable mirrors. *J. Appl. Phys.* **86**, 6107–6119 (1999).
- Huiberts, J. N. *et al.* Synthesis of yttrium trihydride films for ex-situ measurements. *J. Alloys Compounds* **239**, 158–171 (1996).
- Hu, M. S. & Evans, A. G. The cracking and decohesion of thin films on ductile substrates. *Acta Metall.* **37**, 17–925 (1989).

Acknowledgements

We thank B. Dam, D. Nagengast, M. van Gogh, M. Welling, R. Elberse, A. Pundt and P. van der Sluis for support, advice and discussions. This work was supported by the Stichting voor Fundamenteel Onderzoek der Materie (FOM), which is financed by NWO. We also acknowledge financial contribution from the European Commission through the TMR program (research network 'Switchable metal hydride films').

Correspondence and requests for materials should be addressed to R.G. (e-mail: kers@nat.vu.nl).

Electrostatic trapping of ammonia molecules

Hendrick L. Bethlem, Giel Berden, Floris M. H. Crompvoets, Rienk T. Jongma, André J. A. van Roij & Gerard Meijer

FOM-Institute for Plasma Physics Rijnhuizen, PO Box 1207, NL-3430 BE Nieuwegein, The Netherlands, and Department of Molecular and Laser Physics, University of Nijmegen, Toernooiveld 1, NL-6525 ED Nijmegen, The Netherlands

The ability to cool and slow atoms with light for subsequent trapping^{1–3} allows investigations of the properties and interactions of the trapped atoms in unprecedented detail. By contrast, the complex structure of molecules prohibits this type of manipulation, but magnetic trapping of calcium hydride molecules thermalized in ultra-cold buffer gas⁴ and optical trapping of caesium dimers⁵ generated from ultra-cold caesium atoms have been reported. However, these methods depend on the target molecules being paramagnetic or able to form through the association of atoms amenable to laser cooling^{6–8}, respectively, thus restricting the range of species that can be studied. Here we describe the slowing of an adiabatically cooled beam of deuterated ammonia molecules by time-varying inhomogeneous electric fields^{9,10} and subsequent loading into an electrostatic trap. We

are able to trap state-selected ammonia molecules with a density of 10^6 cm^{-3} in a volume of 0.25 cm^3 at temperatures below 0.35 K . We observe pronounced density oscillations caused by the rapid switching of the electric fields during loading of the trap. Our findings illustrate that polar molecules can be efficiently cooled and trapped, thus providing an opportunity to study collisions and collective quantum effects in a wide range of ultra-cold molecular systems^{11–14}.

Adiabatic cooling in an expansion can be used to obtain high densities of cold molecules. In pulsed molecular beams, densities on the order of 10^{12} molecules per cm^3 in a single quantum state at translational temperatures of 1 K are readily achieved. Typical velocities in a molecular beam are in the $250\text{--}2,000 \text{ m s}^{-1}$ range, however, preventing trapping of these dense low-temperature samples in the laboratory frame. Various methods are currently being explored to translate the high phase-space densities from the moving frame of the molecular beam to the laboratory frame^{9,10,15–17}. Once molecules are present at sufficiently low kinetic energies they can be trapped in various ways, for instance, by using the rather deep traps ($\sim 1 \text{ K}$) that can be formed with inhomogeneous magnetic¹⁸ or electric fields¹⁹.

Our approach to slowing down and trapping a beam of neutral molecules exploits the interaction of dipolar molecules with electric fields. The experimental set-up consists of a compact molecular beam machine, schematically indicated in Fig. 1. A pulsed beam of ND_3 is produced by expanding a mixture of less than 1% ND_3 (where D is deuterium) in Xe into vacuum, using a modified solenoid valve. Seeding of ND_3 in the heaviest rare gas, together with cooling of the pulsed valve, results in a mean beam velocity of 280 m s^{-1} (kinetic energy $E_{\text{kin}} = 66 \text{ cm}^{-1}$). The velocity spread is approximately 15%, corresponding to a translational temperature of about 0.8 K in the moving frame. Only molecules in states that gain Stark energy with increasing electric field, that is, molecules in so-called 'low-field seeking' states, are selected in these experiments. The only low-field seeking state populated in the molecular beam is the $|J K |M\rangle = |1 1 1\rangle$ upper inversion level in the vibrational ground state: approximately one-eighth of the total number of ND_3 molecules is in this particular quantum state. The electric dipole moment of ND_3 is 1.5 debye (ref. 20), and molecules in this particular level experience a positive Stark shift of 1.2 cm^{-1} in an electric field of 100 kV cm^{-1} . We selected ND_3 rather than NH_3 for these experiments because the magnitude of the inversion splitting is significantly reduced upon deuteration, resulting in a more linear Stark effect for ND_3 in the applied electric fields²⁰.

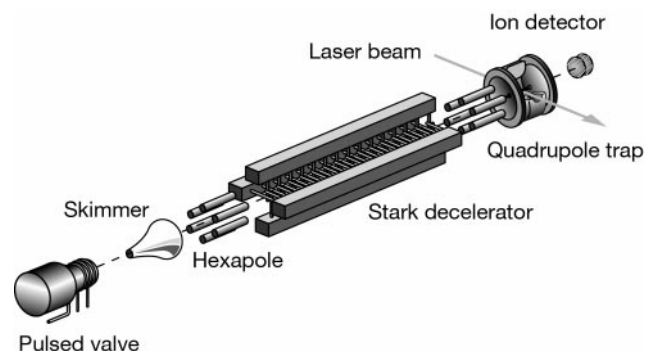


Figure 1 Experimental set-up. From left to right a gas pulse of ND_3 in Xe exits a cooled ($T = 200 \text{ K}$) pulsed valve with a mean velocity of 280 m s^{-1} . After passage through the skimmer the beam is collimated with a pulsed hexapole and a subset of these molecules is subsequently slowed down upon passage through the Stark decelerator. The package of slow molecules is focused into the electrostatic trap. Part of the ring electrode is not drawn to allow a view of the inside of the trap. Molecules at the centre of the trap are ionized with the laser; ions are extracted and detected with an ion detector. The flight path of the molecules from the pulsed valve to the centre of the trap is 56 cm .

The ND_3 molecules pass through a 1.0-mm diameter skimmer into a second, differentially pumped, vacuum chamber and fly into a short pulsed hexapole that acts as a positive lens for the selected ND_3 molecules. After exiting the hexapole the molecules enter a 35-cm long Stark decelerator. The decelerator consists of an array of 63 equidistant electric field stages. When the state-selected ND_3 molecules enter a region of high electric field (typically up to 100 kV cm^{-1}), they will gain Stark energy. This gain in Stark energy ('potential' energy) is compensated by a loss in kinetic energy. If the electric field is greatly reduced before the molecules have left this region they will not regain the lost kinetic energy. This process is repeated by letting the molecules pass through the series of electric field stages, which are switched synchronously with the arrival of the package of decelerating molecules in these stages. This provides a method of decelerating a sample of neutral molecules while maintaining the initial phase-space density^{9,10}, much as is done in charged-particle accelerators.

A second pulsed hexapole at the end of the decelerator focuses the slow molecules exiting the decelerator into the electrostatic trap. The trap has a quadrupole geometry¹⁹ and a cut through the centre of the trap is shown in the upper part of Fig. 2. The 2-mm diameter entrance and exit holes in the end-caps enable the molecular beam to pass through the trap. The inner radius of the ring electrode is 5 mm . There are 2-mm diameter holes in this electrode to allow probing of the density of ND_3 molecules at the centre of the trap via a laser-ionization detection scheme. Using the focused radiation of a pulsed laser at 302.4 nm , only the ND_3 molecules in the selected state are multi-photon ionized²¹. Just before the probe laser is fired, the voltages on the entrance end-cap as well as on the ring electrode are switched off. A constant voltage of -400 V is applied to the exit end-cap. The trap thus serves as the extraction region of a linear time-of-flight mass spectrometer, and the ND_3 -ion signal is recorded using an ion detector placed further downstream.

In order to transport the molecules into the quadrupole trap, some residual velocity is required. In the lower trace of Fig. 3, the arrival-time distribution of decelerated ND_3 molecules at the centre of the trap is shown relative to the time when the pulsed valve is opened. No voltages are applied to the entrance end-cap and ring

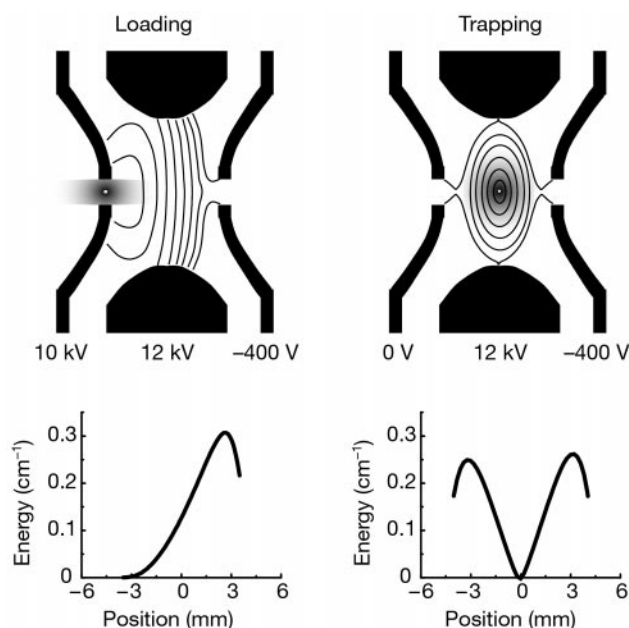


Figure 2 Configuration of the trap with the voltages as applied during loading and trapping. In the trap, lines of equal electric field are indicated and the cloud of molecules is sketched. The potential energy along the molecular beam axis of the ND_3 molecules in the $|J K |M\rangle = |1 1 1\rangle$ state with positive Stark shift is shown for both field geometries.

electrode. In this particular case, the Stark decelerator is set to decrease the kinetic energy of the ND_3 molecules by about 0.9 cm^{-1} per electric field stage, starting from an initial mean velocity of 260 m s^{-1} ($E_{\text{kin}} = 57 \text{ cm}^{-1}$). It should be noted that only 1% of the ND_3 molecules in the molecular beam that are in the right quantum-state are also in the region of phase-space that is accepted by the Stark decelerator with these settings. The mean velocity of the package of slow molecules arriving at the centre of the trap at 7.7 ms is 13 m s^{-1} ($E_{\text{kin}} = 0.13 \text{ cm}^{-1}$), and has a relatively narrow velocity spread of 2 m s^{-1} at full-width half-maximum (FWHM). The latter corresponds to a translational temperature in the moving frame of the order of 2 mK.

To load the trap, the voltages on the trap are applied asymmetrically, as shown in Fig. 2. The ND_3 molecules thus have to climb a last potential hill, and are brought to a standstill near the centre of the trap. The measured density of the ND_3 molecules at the centre of the trap as a function of time in this field geometry is shown in the middle trace of Fig. 3. As expected, the molecules are observed to reach the centre of the trap somewhat later. Basically, the data show the reflection of the slow molecules from the repulsive wall ('concave mirror') of the trap.

To actually trap the molecules, we switch the voltage on the entrance end-cap to ground potential when the molecules have come to a near-standstill inside the trap. We thereby create a (near) symmetric electric field configuration with a minimum at the geometrical centre of the trap. The depth of the potential well in which the ND_3 molecules are then confined is 0.24 cm^{-1} (0.35 K), giving a strict upper limit to the temperature of the sample of trapped molecules. Depending on the details of the trap-loading,

the actual temperature of the molecules in the trap might be as low as the translational temperature originally present in the moving frame of the decelerated molecular beam. The measured density of ND_3 at the centre of the trap as a function of time is shown in the upper trace of Fig. 3. At early times after switching on the trap, clear oscillations in the density with a period of 0.58 ms are observed. These oscillations are a signature of the bouncing of the molecules in the three-dimensional potential well when still only a relatively small region of space is occupied by the trapped molecules. The oscillations fade away after several milliseconds and a steady signal from the trapped ND_3 molecules remains. The actual signal observed from the trapped molecules amounts to a few ND_3 ions per laser pulse. As ionization of the trapped ND_3 molecules only takes place in the focus of the laser beam, and as the resulting ions have to be extracted through the 2-mm diameter hole in the exit end-cap, the actual detection volume is less than 10^{-5} cm^3 . This translates into a lower limit for the density of trapped, state-selected ammonia molecules of 10^6 cm^{-3} .

To determine the trapping lifetime, the ND_3 density at the centre of the trap is monitored as a function of the time elapsed since loading of the trap. The results are displayed in the inset of Fig. 3. An exponential decay of the number of trapped molecules with a $1/e$ decay time of $0.24 \pm 0.04 \text{ s}$ is observed. The measured background pressure in the vacuum chamber is around $2 \times 10^{-8} \text{ mbar}$, and collisions with background-gas are therefore most probably responsible for the observed trap-loss rate. Owing to the increased pressure in the chamber shortly after releasing the gas pulse, the trap loss rate is observed to be slightly higher at early trapping times.

The deceleration and trapping scheme presented here can be applied to a large variety of neutral molecules. For this, the molecule under study must have a sufficiently large positive Stark shift (about 1 cm^{-1}) in experimentally feasible electric fields (about 200 kV cm^{-1}). Furthermore, a pulsed beam with sufficiently low initial kinetic energy is required. On the basis of these criteria, molecules such as OH, $\text{NH}(a^1\Delta)$, NO, $\text{CO}(a^3\Pi)$, H_2O , NO_2 , H_2CO and CH_3F appear to be excellent candidates for future trapping experiments. It is expected that the trapped molecules can then be further cooled using evaporative cooling or using laser-cooling on ro-vibrational transitions. This opens the way for the study of cold molecule-molecule collisions, collective quantum effects in molecular systems, the physics of mutually interacting dipolar molecules and many other applications already realized for atoms^{18,22}.

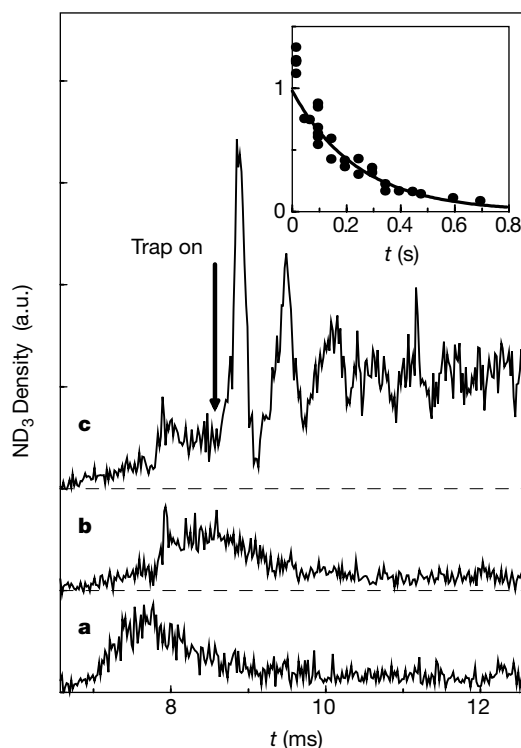


Figure 3 Density of ammonia molecules at the centre of the trap as a function of time. The horizontal axis shows the time after opening of the pulsed valve. **a–c**, The three traces show the ND_3 density at the centre of the trap when no voltages are applied to the entrance end-cap and the ring electrode (**a**), when voltages are applied as depicted for the loading scheme in Fig. 2 (**b**) and when voltages are applied for loading and trapping (**c**). The time that switching from the loading to the trapping scheme occurs is indicated by an arrow. The inset shows the density of trapped ammonia molecules on a longer timescale. The dashed curve shows an exponential curve with a $1/e$ decay time of 0.24 s.

Received 30 May; accepted 30 June 2000.

1. Chu, S. Manipulation of neutral particles. *Rev. Mod. Phys.* **70**, 685–706 (1998).
2. Cohen-Tannoudji, C. N. Manipulating atoms with photons. *Rev. Mod. Phys.* **70**, 707–719 (1998).
3. Phillips, W. D. Laser cooling and trapping of neutral atoms. *Rev. Mod. Phys.* **70**, 721–741 (1998).
4. Weinstein, J. D., de Carvalho, R., Guillet, T., Friedrich, B. & Doyle, J. M. Magnetic trapping of calcium monohydride molecules at millikelvin temperatures. *Nature* **395**, 148–150 (1998).
5. Takekoshi, T., Patterson, B. M. & Knize, R. J. Observation of optically trapped cold cesium molecules. *Phys. Rev. Lett.* **81**, 5105–5108 (1998).
6. Band, Y. B. & Julienne, P. S. Ultracold-molecule production by laser-cooled atom photoassociation. *Phys. Rev. A* **51**, R4317–R4320 (1995).
7. Fioretti, A. *et al.* Formation of cold Cs_2 molecules through photoassociation. *Phys. Rev. Lett.* **80**, 4402–4405 (1998).
8. Nikolov, A. N. *et al.* Observation of ultracold ground-state potassium molecules. *Phys. Rev. Lett.* **82**, 703–706 (1999).
9. Bethlem, H. L., Berden, G. & Meijer, G. Decelerating neutral dipolar molecules. *Phys. Rev. Lett.* **83**, 1558–1561 (1999).
10. Bethlem, H. L., Berden, G., van Roij, A. J. A., Crompvoets, F. M. H. & Meijer, G. Trapping neutral molecules in traveling potential well. *Phys. Rev. Lett.* **84**, 5744–5747 (2000).
11. Doyle, J. M. & Friedrich, B. Molecules are cool. *Nature* **401**, 749–751 (1999).
12. Williams, C. J. & Julienne, P. S. Molecules at rest. *Science* **287**, 986–987 (2000).
13. Wynar, R., Freedland, R. S., Han, D. J., Ryu, C. & Heinzen, D. J. Molecules in a Bose-Einstein condensate. *Science* **287**, 1016–1019 (2000).
14. Herschbach, D. Molecular clouds, clusters, and corrals. *Rev. Mod. Phys.* **71**, S411–S418 (1999).
15. Maddi, J. A., Dinneen, T. P. & Gould, H. Slowing and cooling molecules and neutral atoms by time-varying electric-field gradients. *Phys. Rev. A* **60**, 3882–3891 (1999).
16. Gupta, M. & Herschbach, D. A mechanical means to produce intense beams of slow molecules. *J. Phys. Chem. A* **103**, 10670–10673 (1999).
17. Friedrich, B. Slowing of supersonically cooled atoms and molecules by time-varying nonresonant induced dipole forces. *Phys. Rev. A* **61**, 025403-1/4 (2000).

18. Friedrich, B. *et al.* Towards magnetic trapping of molecules. *J. Chem. Soc. Faraday Trans.* **94**, 1783–1791 (1998).
19. Wing, W. H. Electrostatic trapping of neutral atomic particles. *Phys. Rev. Lett.* **45**, 631–634 (1980).
20. Gandhi, S. R. & Bernstein, R. B. Focusing and state selection of NH_3 and OCS by the electrostatic hexapole via first- and second-order Stark effects. *J. Chem. Phys.* **87**, 6457–6467 (1987).
21. Ashfold, M. N. R., Dixon, R. N., Little, N., Stickland, R. J. & Western, C. M. The $\text{B}^1\Sigma^+$ state of ammonia: sub-Doppler spectroscopy at vacuum ultraviolet energies. *J. Chem. Phys.* **89**, 1754–1761 (1988).
22. Bahns, J. T., Gould, P. L. & Stwalley, W. C. Formation of cold ($T \leq 1$ K) molecules. *Adv. At. Mol. Opt. Phys.* **42**, 171–224 (2000).

Acknowledgements

This work is part of the research program of the 'Stichting voor Fundamenteel Onderzoek der Materie (FOM)', which is financially supported by the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)'. The research of R.T.J. has been made possible by a fellowship of the Royal Netherlands Academy of Arts and Sciences. We acknowledge the technical assistance of Ch. Timmer.

Correspondence and requests for information should be addressed to G.M.
(e-mail: gerardm@rijnh.nl).

Quasi-planar nucleus structure in apoferritin crystallization

S.-T. Yau* & Peter G. Vekilov*†

* Center for Microgravity and Materials Research, and † Department of Chemistry, University of Alabama in Huntsville, Huntsville, Alabama 35899, USA

First-order phase transitions of matter, such as condensation and crystallization, proceed through the formation and subsequent growth of 'critical nuclei' of the new phase. The thermodynamics and kinetics of the formation of these critical nuclei depend on their structure, which is often assumed to be a compact, three-dimensional arrangement of the constituent molecules or atoms^{5,6}. Recent molecular dynamics simulations have predicted compact nucleus structures for matter made up of building blocks with a spherical interaction field^{7,8}, whereas strongly anisotropic, dipolar molecules may form nuclei consisting of single chains of molecules⁹. Here we show, using direct atomic force microscopy observations, that the near-critical-size clusters formed during the crystallization of apoferritin, a quasi-spherical protein, and which are representative of the critical nucleus of this system, consist of planar arrays of one or two monomolecular layers that contain 5–10 rods of up to 7 molecules each. We find that these clusters contain between 20 and 50 molecules each, and that the arrangement of the constituent molecules is identical to that found in apoferritin crystals. We anticipate that similarly unexpected critical nucleus structures may be quite common, particularly with anisotropic molecules, suggesting that advanced nucleation theories should treat the critical nucleus structure as a variable.

Protein crystallization is a convenient model for studies of molecular structures and dynamics of nucleation: protein molecules' sizes (a few nanometres) and the typical timescales for growth (a few seconds between sequential discrete growth events) are within the reach of current surface characterization techniques. The molecules of our model, apoferritin¹⁰, are quasi-spherical and consist of 24 subunits arranged in pairs along the 12 walls of a quasi-rhombododecahedron^{11,12}. Apoferritin crystals have a face-centred cubic (f.c.c.) lattice, faceted by hexagonal [111] planes^{13,14}.

We monitored large crystals for periods of several hours in the range of supersaturations σ from 0.5 to 1.6 (for experimental procedures and definition of σ , see Methods below), and detected clusters landing on the top (111) crystal surfaces. Figures 1a and 2a,

taken a few minutes after cluster landing, show examples representative of more than 15 such events. In all cases, the molecules in a cluster are arranged in rows of 4–8 molecules, while 3–7 rows assemble in planar domains with an additional 2–3 rows forming a second layer. The cluster in Fig. 1 contains two domains linked by a longer row of about 10 molecules. The centre-to-centre distance between adjacent molecules in a row is 13 nm, equal to that along the close-packed (110) direction in the crystal. Furthermore, one of the molecular rows of the cluster in Fig. 2 generates a new (111) layer that spreads on the crystal surface to meet the crystal's own layer. Hence, these are clusters of apoferritin molecules, which, unlike occurrences in other systems^{15,16}, have the same arrangement as in the crystal.

The quasi-planar cluster shape precludes the clusters being pieces chopped off a large crystal. We conclude that the clusters form in the solution bulk and then land on the monitored surface because: (1) the clusters consist of (110) planes (Fig. 1f) rather than (111) planes, typical of 'two-dimensional' nuclei of new layers replicating the (111) molecular arrangement of the crystal surface¹⁴. (2) The molecular rows in the cluster in Fig. 1 are at an angle to the crystal's (110) direction. (3) The cluster and the layer originating from it in Fig. 2 are out of registry with the crystal's own layer causing a boundary free of molecules or consisting of strained molecules. (4) Often—see for instance Fig. 1e—the clusters are pushed back into the solution by the advancing crystal layers. (5) If trapped by the crystal layers¹⁷, as in Fig. 2e and f, a misfit boundary between the two structures appears. Did the cluster structure change after landing, under the influence of the translationally symmetric force exerted by the underlying crystal? It seems not: we would expect that in such a case there would be an exact match and continuity between the structures of the cluster and the underlying crystal, as observed before^{15,18}.

Figures 1 and 2 show that molecules attach to, and detach from, the cluster between two frames. The attachment and detachment frequencies are comparable, which is unusual for the supersaturated conditions of the observation (we note the fast advancing long straight steps in Fig. 1d and e, and Fig. 2e and f). Comparable rates of molecular attachment and detachment, and bifurcation of their subsequent evolution into either growth or dissolution were observed for all clusters of such sizes seen in our experiments. Estimating the average net frequency of molecular attachment in Fig. 1, we get about 4 molecules/43 s ≈ 0.1 s⁻¹ for the approximately 20 possible attachment sites at the ends of the molecular rows, or 0.005 s⁻¹ per attachment site. This frequency is more than an order of magnitude lower than the net frequency of attachment to a growth site on the surface of a large crystal—0.065 s⁻¹ (ref. 14). The ratio of the two rates indicates that the size of the cluster in Fig. 1 is just above the critical size. The cluster in Fig. 2 loses one layer of 6 molecules in 896 s (compare Fig. 2a and d). Hence, its size must be just below the critical for that supersaturation. We conclude that these are near-critical-size clusters for the phase transformation occurring in the system, crystallization of apoferritin. Their sizes are, respectively, larger and smaller than the critical, and their structure should be representative of the structure of the nucleus.

The size of the near-critical cluster in Fig. 2 at $\sigma = 1.6$ decreases from about 25 to about 20 molecules during the monitoring time. This cluster is smaller than the one in Fig. 1 at $\sigma = 1.1$ which contains about 50–60 molecules. On the average, smaller near-critical clusters were observed at higher supersaturations. Although we do not have sufficient statistics for a quantitative statement, these observations agree with the predictions of the classical and advanced treatments of nucleation^{1,2,19}.

To eliminate possible effects of the large crystals on the nucleation pathways, we viewed the glass bottom of the atomic force microscope (AFM) cell before the formation of any crystals at σ between 0.5 and about 2.5. A disordered apoferritin layer with roughly hexagonal molecular co-ordination covers the glass. We saw numer-

ous clusters ranging from 2–3 to 50–70 molecules (we never saw single molecules) that land on the apoferritin-covered glass, stay adsorbed for 1 to 30 min and then dissolve or desorb. Typically, smaller clusters had significantly shorter residence times on the surface. Hence, although the occurrence of smaller clusters appeared higher, comparisons of the populations of large and

small clusters with theoretical predictions would be unjustified. We noted that: (1) in the smallest clusters the molecules occupy the corners of a polygon. (2) Clusters of about 10–20 molecules consist of two parallel molecular rows. (3) Two structures were possible for clusters of 20 or more molecules: a few molecular rows in a plane, or disordered aggregates. (4) Apoferritin molecules detached from the

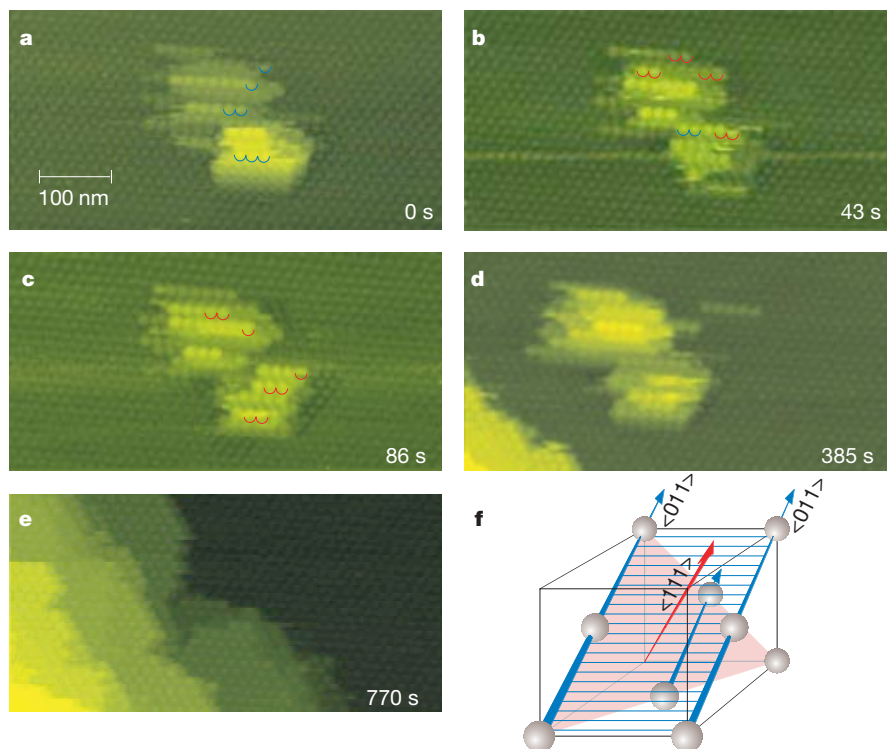


Figure 1 Near-critical-size cluster on the (111) face of an apoferritin crystal at supersaturation $\sigma = 1.1$. **a–c**, Molecules attach and detach from the cluster. Molecules that are missing in the next frame are highlighted in blue, those that have appeared after the previous frame was captured are highlighted in red. **d, e**, An advancing step pushes the cluster back into the solution. The step velocity is 0.6 nm s^{-1} , close to the fastest step

velocity previously recorded for this supersaturation with apoferritin crystals¹⁴. This indicates that the selected AFM imaging mode did not affect the monitored processes. **f**, Diagram of (111) and (110) planes and <110> molecular rows in a face-centred cubic (f.c.c.) crystal lattice.

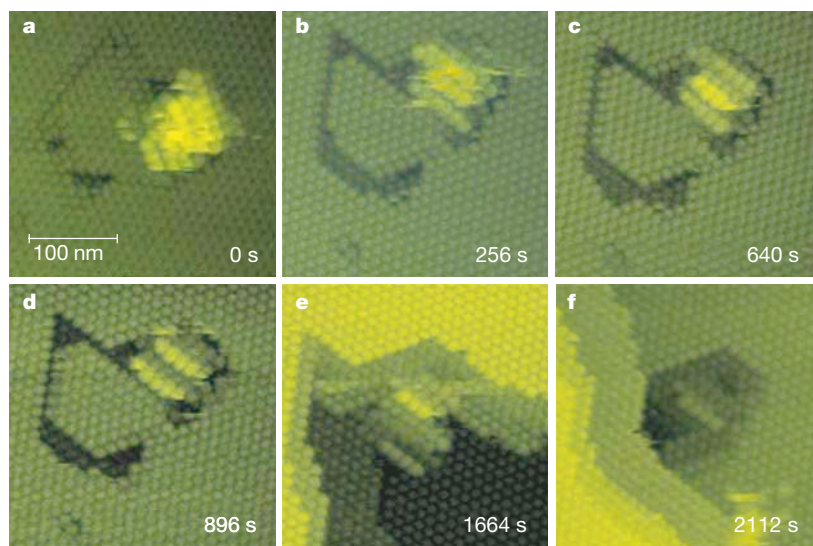


Figure 2 A slowly dissolving cluster at $\sigma = 1.6$. **a–d**, The top layer of the cluster is gradually disappearing because more molecules detach than attach to the cluster. The fraction of the top crystal layer originating from the cluster is out of registry with the remaining part. **e, f**, A few steps belonging to the crystal surround the cluster. Note the

misfit between the molecules of the crystal and the molecules of the cluster. The direction of scanning was changed between **a** and **b**. The consistency of the cluster structure is evidence for the lack of AFM imaging artefacts.

ordered clusters, while (5) the disordered ones were inactive. (6) The larger ordered clusters were not seen at the high supersaturations. We conclude that the clusters in (1)–(4) above are the subcritical clusters that constitute the stages in the nucleation process preceding those seen in Figs 1 and 2.

Figure 3 shows that the clusters that form in the solution may develop into (111) faceted f.c.c. crystals. We see a microcrystal that consists of three (111) layers of about 60 molecules in each. The inclination of the microcrystals indicates that it was formed in the solution bulk and later landed in our field of view. Clusters similar to those in Figs 1 and 2 can evolve to such a microcrystal by accumulating several (110) layers until a (111) face is formed.

The nucleation pathway emerging from the above observations is schematically summarized in Fig. 4. The (110) planar—rather than compact as in Fig. 4b—structure of the nucleus is surprising^{5,6}. At present, in analogy to the case of anisotropic dipolar molecules⁹, we can only speculate that the slight anisotropy of the apoferritin molecules, either directly, or after enhancement by the formation of a two-member cluster, underlies the observed shape.

This nucleus structure may have important consequences for the nucleation process. A planar cluster has a larger surface area than a compact cluster with the same number of molecules n . As a result, a larger contribution to the crystallization energy gain is needed to compensate for the greater surface energy loss, and hence the number of molecules in the critical cluster n^* is greater. Because the nucleation barrier $\Delta G^*(n^*) \approx n^* \Delta \mu / 2$ (refs 5, 19–21), the higher nucleation barriers lead to slower nucleation kinetics than predicted by the classical theories based on compact spherical nuclei. Furthermore, the rough surface of the nuclei in Figs 1–3 may result in a surface energy that is not a smooth and monotonic function of n . This may result in unusual dependencies of the nucleation rate on $\Delta \mu$.

To quantitatively characterize the nucleation process, we use the fact (see Figs 3 and 4a), that the nucleus evolves into a crystal by accumulating new layers. Following refs 22 and 23, we find that the contribution to the partial molecular surface free energy associated

with the equilibrium addition of a (110) layer consisting of $n_e \times n_e$ molecules to a (110) cluster surface is $2\phi/n_e$. Here ϕ is the free energy of the intermolecular bond in the crystals (the entropic components stem from the water molecules trapped in the crystal or bound to the solute; S.-T.Y. *et al.*, manuscript in preparation). For a critical cluster in labile equilibrium with the solution, this excess free energy is balanced by the supersaturation, that is, $\Delta \mu = 2\phi/n_e$ (the Gibbs–Thomson equation^{22,23}). Substituting $\phi/k_B T = 3.2$ (ref. 14), with $\Delta \mu/k_B T = 1.1$ and 1.6, the values of n_e are 6 and 4, respectively, that is, approximately the cluster sizes in Figs 1 and 2. Hence, ϕ is the molecular-level equivalent to surface tension that governs nucleation²².

Comparing the n_e values with the respective n^* values indicated above, we find that $n^* \approx n_e^{2.3}$. Extrapolating to $\Delta \mu/k_B T = 3.8$, at which apoferritin crystals are typically grown¹³, we get $n_e \approx 1$ or 2, $n^* \approx 3$ –4 (for structures of such clusters imaged by atomic force microscopy at lower σ values, see above). Assuming a pre-exponential factor for the nucleation rate law $J_0 \approx 1 \text{ cm}^{-3} \text{ s}^{-1}$ (for lysozyme, J_0 is between 1 and $10 \text{ cm}^{-3} \text{ s}^{-1}$; refs 24, 25) and using, as above, $\Delta G^* = n^* \Delta \mu / 2$, we get for the nucleation rate $J = J_0 \exp[-\Delta G^*(n^*)/k_B T] \approx 10^{-3} \text{ cm}^{-3} \text{ s}^{-1}$. In an overnight experiment with a few hundred microlitres of solution, about 10 crystals should nucleate. Accordingly, numerous experiments under these conditions produced between a few and about 100 crystals.

To compare the estimated ϕ to previous results, we define a corresponding effective macroscopic surface energy γ as $n_{\text{free}} \phi / S$. Here n_{free} is the number of unsaturated bonds of a molecule on the surface of a nucleus, and S is the surface area of a molecule. With n_{free} being on average 7–8 for the clusters that expose two sides of many molecules, (see Figs 1 and 2) we get $\gamma \approx 0.2 \text{ mJ m}^{-2}$. Light-scattering determinations of apoferritin nuclei sizes averaged over all body angles²⁶ at $\sigma = 0.92$ yielded values of about 40 nm (or ~ 3.5 molecular dimensions, compatible with the cluster in Fig. 1). Assuming a spherical nucleus shape, a value of $\gamma = 0.027 \text{ mJ m}^{-2}$ was obtained²⁶. As discussed above, for a spherical cluster n_{free} is lower than for a flat cluster with the same n_e . Hence, the surface energy value resulting from this assumption is also lower. Similarly, AFM studies of virus crystallization kinetics²⁷ produced γ -values

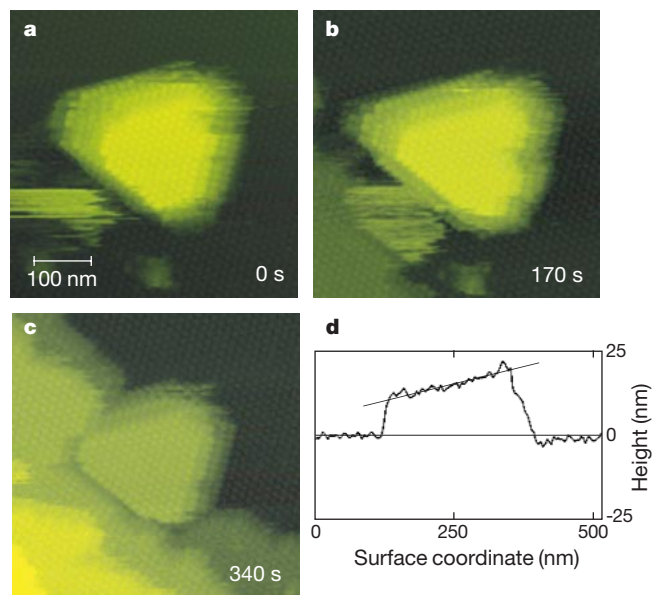


Figure 3 Microcrystal at $\sigma = 1.1$. **a–c**, The microcrystal slowly grows by attachment of molecules to the side (110) faces, while a step belonging to the crystal moves rapidly to incorporate it. **d**, Height profile along a line roughly from top left to bottom right in **a** shows that the microcrystal is inclined with respect to the substrate by about 3° . This corresponds to one molecule trapped under the microcrystal close to its edge or to an unfinished layer on its bottom surface.

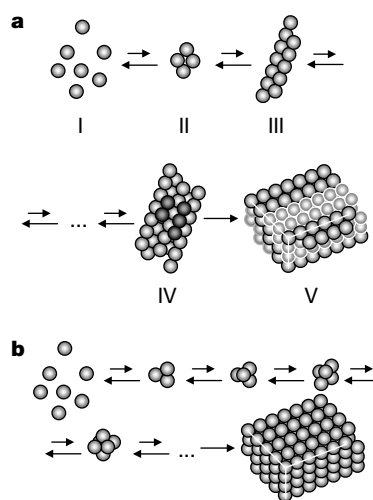


Figure 4 Schematic illustration of nucleation pathways. **a**, Nucleation via a planar critical cluster. Stage I, single molecules in the solution; the lack of any significant concentration of dimers, trimers, and so on, has been shown by static and dynamic light scattering^{28,29}. Stage II, a few molecules at the corners of a polygon. Stage III, linear array. Stage IV, a quasi-planar critical cluster with (110) orientation, similar to structures seen in Figs 1 and 2; molecules belonging to the second layer are shown in a lighter shade. Stage V, microcrystal faceted by (111) planes; the (110) layers that stack up to form this crystal are delineated by lighter and darker contours. **b**, Nucleation via a compact critical cluster^{5,6}.

higher by about the same factor than the evaluation based on light scattering²⁶. □

Methods

Monitoring was performed *in situ* in the AFM fluid cell at $23.0 \pm 0.3^\circ\text{C}$ maintained by stabilizing the room temperature. This temperature was higher by $0.5\text{--}1^\circ\text{C}$ than the setting in the room. The insensitivity of apoferritin crystallization to temperature variations¹³ justifies this approach to temperature control. We used the less intrusive tapping mode of the AFM and a tip with minimal force constant. The scanning parameters were adjusted such that continuous imaging affected neither the surface structure, nor the process dynamics. For verification, we varied the scan sizes and the time elapsed between image collections, and saw that neither the spatial nor the temporal characteristics of the monitored processes changed. Supersaturation σ , defined as chemical potential difference $\Delta\mu$ (in $k_B T$ units) between the solution and a large crystal, was calculated from the actual and equilibrium solution concentrations, C and C_c : $\sigma \equiv \Delta\mu/k_B T = \ln(C/C_c)$. The effects on solution non-ideality on this determination of $\Delta\mu$ are minor for this system (S.-T.Y. *et al.*, manuscript in preparation). This $C_c = 23 \pm 3 \mu\text{g cm}^{-3}$ was determined as the value of C at which long steps stopped moving, before retreating in solutions of $C < C_c$. The (111) surface provides a scale for determinations of the dimensions of the clusters. The periodicity within a molecular row along a $\langle 110 \rangle$ direction is 13 nm, and the layer thickness is 10.5 nm, in good agreement with the X-ray structure^{11,12}.

AFM, a surface characterization technique, can be applied to visualize clusters that appear in the solution bulk only because they reach a surface and adsorb on it. The average brownian diffusion time τ to reach the substrate for a cluster formed within a $x = 100 \mu\text{m}$ thick layer can be evaluated from Einstein's relation $x^2 = 2D\tau$. A lower estimate for cluster diffusivity D can be obtained from the diffusivity of single apoferritin molecules, $3.2 \times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$ (refs 28, 29), using Stokes law and assuming that the cluster behaves like a particle with seven molecules at an edge: $D \approx 5 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$. Substituting, $\tau \approx 1,000 \text{ s} \approx 15 \text{ min}$. Assuming sedimentation in the Earth's gravity field³⁰ would lead to landing times longer by more than an order of magnitude.

We performed numerical simulations of combined buoyancy-driven convection and solute diffusion, as in ref. 31, to evaluate the concentration nonuniformity due to the crystal growth in the cell. We found that at the average rate of growth of the underlying crystal $< 0.1 \text{ nm s}^{-1}$, apoferritin depletion at the interface is $< 1\%$, and the characteristic diffusion and convection velocities are, respectively, $0.3 \mu\text{m s}^{-1}$ and $< 1 \mu\text{m s}^{-1}$. The AFM tip travel for scan widths of $0.5 \mu\text{m}$ and frequencies of about 3 s^{-1} , and the $\sim 20\text{-kHz}$ tapping oscillation, could add solution flow with similar velocities. None of those should affect cluster evolution.

Received 3 September 1999; accepted 8 June 2000.

- Gibbs, J. W. *The Collected Works of J. W. Gibbs* (Yale Univ. Press, New Haven, 1961).
- Oxtoby, D. W. Nucleation of first order phase transitions. *Acc. Chem. Res.* **31**, 91–97 (1998).
- Volmer, M. *Kinetik der Phasenbildung* (Steinkopff, Dresden, 1939).
- Hill, T. L. *Thermodynamics of Small Systems* (Benjamin, New York, 1963).
- Milchev, A. Electrochemical phase formation on a foreign substrate - basic theoretical concepts and some experimental results. *Contemp. Phys.* **32**, 321–332 (1991).
- Chernov, A. A. *Modern Crystallography III: Growth of Crystals* (Springer, Berlin, 1984).
- ten Wolde, P. R. & Frenkel, D. Enhancement of protein crystal nucleation by critical density fluctuations. *Science* **277**, 1975–1978 (1997).
- Talanquer, V. & Oxtoby, D. W. Crystal nucleation in the presence of a metastable critical point. *J. Chem. Phys.* **109**, 223–227 (1998).
- ten Wolde, P. R., Oxtoby, D. W. & Frenkel, D. Coil-globule transition in gas-liquid nucleation of polar fluids. *Phys. Rev. Lett.* **81**, 3695–3698 (1998).
- Harrison, P. M. & Arosio, P. The ferritins: molecular properties, iron storage function and cellular regulation. *Biochim. Biophys. Acta* **1275**, 161–203 (1996).
- Lawson, D. M. *et al.* Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature* **349**, 541–544 (1991).
- Hempstead, P. D. *et al.* Comparison of the three dimensional structures of recombinant human H and horse L ferritins at high resolution. *J. Mol. Biol.* **268**, 424–448 (1997).
- Thomas, B. R., Carter, D. & Rosenberger, F. Effects of microheterogeneity on horse spleen apoferritin crystallization. *J. Cryst. Growth* **187**, 499–510 (1997).
- Yau, S.-T., Thomas, B. R. & Vekilov, P. G. Molecular mechanisms of crystallisation and defect formation. *Phys. Rev. Lett.* **85**, 353–356 (2000).
- Kuznetsov, Y. G., Malkin, A. J. & McPherson, A. Atomic force microscopy studies of phase separations in macromolecular systems. *Phys. Rev. B* **58**, 6097–6103 (1998).
- Georgalis, Y., Umbach, P., Raptis, J. & Saenger, W. Lysozyme aggregation studied by light scattering. I. Influence of concentration and nature of electrolyte. *Acta Crystallogr. D* **53**, 691–702 (1997).
- Malkin, A. J., Kuznetsov, Y. G. & McPherson, A. Defect structure of macromolecular crystals. *J. Struct. Biol.* **117**, 124–137 (1996); Incorporation of microcrystals by growing protein and virus crystals. *Proteins Struct. Funct. Genet.* **24**, 247–252 (1996).
- Kuznetsov, Y. G., Malkin, A. J. & McPherson, A. AFM studies of the nucleation and growth mechanisms of macromolecular crystals. *J. Cryst. Growth* **196**, 489–502 (1999).
- Mutsaers, B. in *Handbook of Crystal Growth* (ed. Hurler, D. T. J.) 189–247 (Elsevier, Amsterdam, 1993).
- Kashchiev, D. On the relation between nucleation work, nucleus size, and nucleation rate. *J. Chem. Phys.* **76**, 5098–5102 (1982).
- Oxtoby, D. W. & Kashchiev, D. A general relation between the nucleation work and the size of the nucleus in multicomponent nucleation. *J. Chem. Phys.* **100**, 7665–7671 (1994).
- Stranski, I. N. & Kaischew, R. Über den Mechanismus des Gleichgewichtes kleiner Kriställchen. I. *Z. Phys. Chem. B* **26**, 100–113 (1934).

- Kaischew, R. & Stranski, I. N. Über die Thomson-Gibbs'sche Gleichung bei Kristallen. *Z. Phys. Chem. B* **35**, 427–432 (1937).
- Vekilov, P. G., Monaco, L. A., Thomas, B. R., Stojanoff, V. & Rosenberger, F. Repartitioning of NaCl and protein impurities in lysozyme crystallization. *Acta Crystallogr. D* **52**, 785–798 (1996).
- Galkin, O. & Vekilov, P. G. Direct determination of the nucleation rate of protein crystals. *J. Phys. Chem.* **103**, 10965–10971 (1999); Are nucleation kinetics of protein crystals similar to those of liquid droplets? *J. Am. Chem. Soc.* **122**, 156–163 (2000); Control of protein crystal nucleation around the metastable liquid-liquid phase boundary. *Proc. Natl Acad. Sci. USA* **97**, 6277–6281 (2000).
- Malkin, A. J. & McPherson, A. Light scattering investigation of the nucleation processes and kinetics of crystallization in macromolecular systems. *Acta Crystallogr. D* **50**, 385–395 (1994).
- Malkin, A. J., Land, T. A., Kuznetsov, Yu. G., McPherson, A. & DeYoreo, J. J. Investigation of virus crystal growth mechanism by *in situ* atomic force microscopy. *Phys. Rev. Lett.* **75**, 2778–2781 (1995).
- Petsev, D. N. & Vekilov, P. G. Evidence for non-DLVO hydration interactions in solutions of the protein apoferritin. *Phys. Rev. Lett.* **84**, 1339–1342 (2000).
- Petsev, D. N., Thomas, B. R., Yau, S.-T. & Vekilov, P. G. Interactions and aggregation of apoferritin molecules in solution: effects of added electrolytes. *Biophys. J.* **78**, 2060–2069 (2000).
- Atkins, P. *Physical Chemistry* (Freeman, New York, 1998).
- Lin, H., Rosenberger, F., Alexander, J. I. D. & Nadarajah, A. Convective-diffusive transport in protein crystal growth. *J. Cryst. Growth* **151**, 153–162 (1995).

Acknowledgements

We thank D. W. Oxtoby, D. N. Petsev, J. I. D. Alexander, O. Galkin, S. Weinkauf, J. M. Garcia-Ruiz and N. Booth for suggestions and encouragement; B. R. Thomas for providing pure apoferritin; H. Lin and D. N. Petsev for numerical simulations; and L. Carver for graphics work. This research was supported by the National Heart, Lung, and Blood Institute, NIH, the Life and Microgravity Sciences and Applications Division of NASA, and the State of Alabama through the Centre for Microgravity and Materials Research at the University of Alabama in Huntsville.

Correspondence and requests for materials should be addressed to P.G.V. (e-mail: peter@cmmr.uah.edu).

Marine control of biological production in the eastern equatorial Pacific Ocean

Paul Loubere

Department of Geology and Environmental Geosciences,
Northern Illinois University, DeKalb, Illinois 60115, USA

The eastern equatorial Pacific Ocean is the site of approximately 20–50% of new biological production in the global oceans¹. This region is also responsible for the greatest efflux of CO_2 from oceans to the atmosphere². New production, which fixes carbon in response to external inputs of nutrients as opposed to supply from local nutrient recycling, is thought to modulate the CO_2 release³. But what controls new production in this region is less clear. Here we present a quantitative reconstruction of biological production in the surface ocean for this region over the past 130,000 years, which shows that the equatorial Pacific Ocean exhibits higher-frequency variations than the South Equatorial Current. Comparison of these records with palaeotemperature reconstructions indicates that atmospherically driven mechanisms—such as aeolian flux of iron or wind-driven changes in upwelling rate of nutrient-rich waters—are unlikely to have influenced longer-term rates of production in this region. Instead, biological production appears to be governed by changes in ocean circulation and the chemical composition of upwelled water.

Four core locations in the eastern equatorial Pacific (EEP) were selected for palaeoproductivity reconstruction with the main objectives of discovering the regional mode(s) of productivity variation and testing for processes controlling that variation. The deep-sea cores used were (Fig. 1): RC13-110 (0.10°N , 95.65°W ; 3,231 m water depth (w.d.); average sedimentation rate (a.s.r.), 2.4 cm kyr^{-1} ,

sampled at 5-cm intervals), Y69-71 (0.10° N, 86.48° W; 2,740 m w.d.; a.s.r., 6.6 cm kyr⁻¹, sampled at 10-cm intervals), ODP 846B (3.095° S, 90.82° W; 3,307 m w.d.; a.s.r., 4.3 cm kyr⁻¹, sampled at 10-cm intervals), and V19-28 (2.47° S, 84.65° W; 2,670 m w.d.; a.s.r., 5 cm kyr⁻¹, sampled at 10-cm intervals). The first two cores lie along the Equator and are subject to shallow equatorial divergence. The last two cores lie below the South Equatorial Current (SEC), which is supplied with nutrients by the Peru upwelling system. The ultimate source of these nutrients appears to be the subantarctic southwest Pacific, and they reach the EEP via the Equatorial Undercurrent⁴. The chronology for the cores is based on previously developed oxygen-isotope stratigraphies of benthic and planktonic foraminifera, which were corrected to calendar years for the interval from 0 to 30,000 years before present⁵⁻⁷.

Quantitative palaeoproductivity estimates were made using the benthic foraminiferal transfer function given in ref. 8. This function is based on the relative abundances of benthic foraminiferal species recovered from the deep-sea sediments, and was created using a global ocean database so as to isolate the benthic foraminiferal response to biological productivity of the surface ocean. The latter exerts a primary control on the flux of organic matter to the sea bed⁹, thus linking the surface ocean to the sea-bed biological community. Factor analysis of surface sediment assemblages in the Pacific and Atlantic oceans shows that benthic patterns clearly reflect modern surface ocean productivity^{10,11}. The transfer function was made by regression of benthic foraminiferal taxon abundances against estimates of annual average surface ocean productivity above each sea-bed sampling location. Surface ocean productivity is not well known, so information from a variety of sources was combined for our benthic sampling locations to produce an internally consistent data set for calibration (see Methods). The r^2 value of the transfer function is 0.89; this is, essentially, an indication that the transfer function is successful in reproducing the relative changes and trends in surface ocean bioproductivity. For the results presented below, the reconstruction of the relative changes in productivity through time is more important than estimation of absolute values.

The palaeoproductivity records are shown in Fig. 2. The cores have a clear pairing of productivity signals, with RC13-110 and Y69-71 along the Equator showing a relatively high-frequency signal which is accentuated during glacial stages 2 and 3. Cores ODP846B and V19-28 both lie below the SEC, and their records are also

strongly similar to one another. But in this case, the records show a lower-frequency pattern and are variable through the entire time span sampled. Each core pairing includes a deeper and a shallower member, and one site with a lower sedimentation rate (hence lower resolution due to bioturbation) and one with a higher rate. The sites correlate by oceanic region, and not by geographic proximity, water depth or sedimentation characteristics. Hence, the benthic foraminiferal transfer function clearly reflects surface ocean properties, and is not affected by deep water mass or sea-bed properties.

The differences between equatorial and SEC productivity signals can be demonstrated quantitatively with principal components analysis of the data. For this, values for the four cores at regular 2,000-year intervals were interpolated from the actual measurements. Interpolation yielded about the same number of data points as in the original records. Principal components are uncorrelated patterns that summarize data variance, so that the first component incorporates the largest portion of that variance and so on. The first principal component (PC1) represents 45% of the data variance, and is shown in Fig. 2. PC1 accounts for 70% of the variance for ODP846 and 67% for V19-28, but only 37% for Y69-71 and 30% for RC13-110. Thus, it incorporates the dominant signal for the SEC core locations, and this signal is of much less importance to the equatorial records. The palaeoproductivity record of the SEC, fed by the Peru upwelling system, is clearly different from that of the equatorial divergence.

The benthic foraminiferal reconstructions in Fig. 2 show generally lower EEP palaeoproductivity during full glacials (isotope stages 2 and 4). For isotope stage 2, this is contrary to estimates based on analyses of organic carbon extracted from the deep-sea sediments. Independent tests supporting the benthic reconstruction, and reasons for the difference from organic carbon estimators are given in ref. 12, where previous reconstructions of equatorial Pacific late Pleistocene oceanography are reviewed. At site V19-28 there is a detailed record of organic carbon accumulation rates¹³. This is not directly comparable to the benthic records presented here as accumulation rates are based on sedimentation rates that must be averaged over variable, and often relatively long, intervals of time; so resolution of changes within the isotopic climate stages is difficult. However, there are similarities. The organic carbon record shows higher values near 125, 105, 75 and 40 kyr, as does the foraminiferal record. The chief difference occurs in the early part of isotope stage 2; here the organic carbon signal increases by a factor of 5 to 10, showing variation much greater than in the record downcore. As no such marked change occurs in the faunal record, it is likely that the organic carbon record is affected by some factor other than surface ocean productivity.

Sea surface temperatures (SSTs) have been reconstructed for the eastern equatorial Pacific using a transfer function based on *Radiolaria* (ref. 14). These reconstructions showed differences in equatorial and SEC patterns of variation, as seen in the benthic records we report here. Some of the data from ref. 14 is presented in Fig. 2. Both the temperature and palaeoproductivity estimates show that the EEP has not behaved uniformly in response to climate factors, but is divided into regions with different response modes. The regionality in the palaeoproductivity data implies that wind-transported materials, such as iron compounds¹⁵, have not been a primary control on EEP surface ocean biological activity. The surface wind vectors estimated by modelling the glacial to interglacial atmosphere¹⁶ make highly unlikely a sharp difference in aeolian flux over the relatively small latitudinal separation between our SEC and equatorial sites. Also, some evidence suggests that aeolian flux did not vary in the necessary way over the late Pleistocene¹⁷.

Traditionally, glacial to interglacial variations in EEP productivity have been linked to supposed changes in equatorial upwelling rate caused by changes in trade-wind velocities. During glacials, wind speeds are thought to have been higher, and increased upwelling

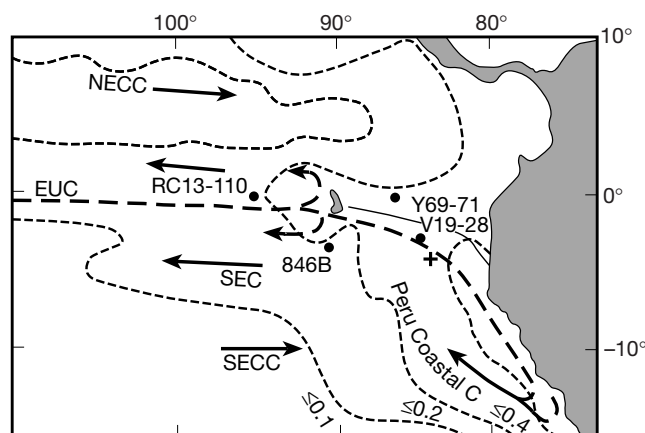


Figure 1 The eastern equatorial Pacific Ocean, showing the locations of the cores examined. The location of V19-29 is shown by the '+' near V19-28. Arrows show the main currents, and the heavy dashed line represents the subsurface Equatorial Undercurrent (EUC). SEC is the South Equatorial Current, and SECC and NECC are the South and North Equatorial Countercurrents. Finer dashed lines enclose regions of phytoplankton concentration (mg m^{-3}) determined by the Coastal Zone Color Scanner.

resulting from this should cool surface waters (SSTs) and increase biological productivity. However, there is a dissenting view that would have equatorial cooling during ice ages linked to advection of colder waters through the eastern boundary currents¹⁸. If upwelling was the dominant process controlling SST and biological productivity, then we would expect to see a good correlation between palaeotemperature and productivity estimates. However, direct comparison of the palaeotemperature and productivity records shown in Fig. 2 reveals that the two do not match. These results indicate that a simple model relating SST and the flux of nutrients, which would control productivity, to upwelling rate cannot explain the observations on a glacial to interglacial timescale. Further, the new palaeoproductivity proxy used here indicates that productivity was lower during full glacials; this is opposite to what is expected if increased trade-wind velocities during colder time periods were driving more upwelling. Finally, a recent model of Pacific atmospheric circulation during the last glacial that incorporates new, lower, SSTs for the region shows little predicted change in the velocity of the zonal winds for the EEP¹⁹, so the ice-age trade winds may not have been much different from the present.

However variable the trade winds and associated upwelling may have been, there are other factors required to account for the glacial–interglacial palaeoproductivity changes; and these factors are regionally distinctive. The results presented here imply that the dominant control on equatorial bioproductivity has been variations in the chemical composition of upwelled water and/or shifts in plankton community composition and ecology. These controls are probably related. Regionality in the EEP palaeoproductivity records may reflect variations in the major and trace nutrient chemistry of the Equatorial Undercurrent (EUC), which has been shown to

consist of several layered flows, the lowest of which emanates from the southwest Pacific Antarctic⁴. The Antarctic region supplying this water lies in a belt between 48° and 55° S, which was strongly affected by glacial–interglacial change. Reconstruction of nutrients for the subantarctic region in the Atlantic and Indian oceans²⁰, which is analogous to that supplying the EUC, indicates that nitrate concentrations and water column stratification differed from the present during the last glacial. If the same was true in the Pacific, then both physical and chemical properties of the EUC were variable on the glacial timescale. Changes in the amount of deep-sea nitrate passed into the South Pacific thermocline could have a dramatic effect on the equatorial system downstream²¹.

Trace nutrient concentrations may also play a role in the longer term bioproductivity record of the EEP. The SEC may be currently iron-limited²², and the EUC can be a source of that element²³. If the iron supply varied on the glacial–interglacial timescale, then it would have an effect on the productivity records we reconstructed. In this context, Farrell *et al.*²⁴ have interpreted nitrogen-isotope data from the SEC region of the EEP to indicate lower nitrate utilization during the last full glacial.

Alternatively, it has been proposed that oceanic silica concentration regulates new production via the activity of diatoms²⁵. For the North Pacific, Berger and colleagues²⁶ have postulated that silica is supplied by deep upwelling and that this supply was reduced during glacials as the large-scale ocean circulation adjusted to the diminished production of North Atlantic Deep Water. Hence, productivity variations might result from a redistribution of silica in the thermocline and changes in the silica/nitrogen ratio of these waters.

The palaeoproductivity data presented here provide a test of the

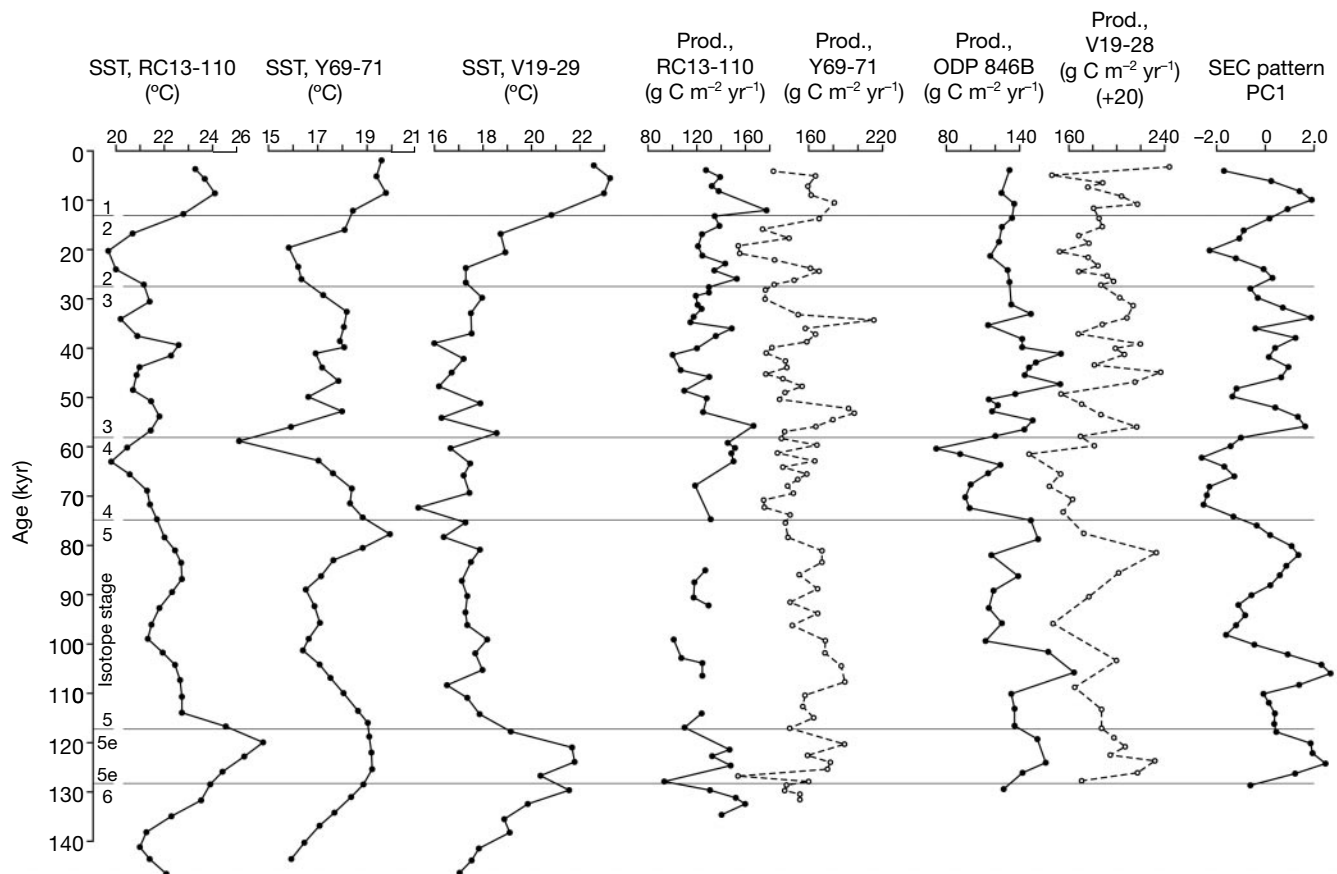


Figure 2 Palaeoproductivity and palaeotemperature records from the eastern equatorial Pacific Ocean. Sea surface temperature (SST) estimates are redrawn from ref. 14. Isotope stages 2, 4 and 6 are glacials. SEC pattern PC1 is the first principal component extracted

from the palaeoproductivity records of the four cores. It correlates strongly to SEC cores ODP846B and V19-28.

hypothesis that a simple physical process, mediated by the trade winds, controls the broad pattern of biological productivity over the eastern equatorial Pacific on the glacial to interglacial timescale. The results do not support this hypothesis. Rather, changing upper-ocean chemistry and circulation appear to have a dominant role and need to be incorporated in a complete model of the EEP. □

Methods

Productivity estimates for calibration with the benthic foraminiferal surface sediment taxon abundance data were developed independently for the Atlantic, Pacific and Indian oceans^{8,11,27,28}. Relative differences and trends in productivity among our benthic sampling locations are primarily based on CZCS (Coastal Zone Color Scanner) contours. Absolute productivity values have been selected to conform with published syntheses of measured productivity, sediment trap organic carbon fluxes, benthic chamber flux data and, where other data are lacking, the Berger²⁹ 'synthetic map' based primarily on oceanic phosphate concentrations. In most cases for our sample set the various sources are within 20% of the Berger²⁹ synthesis of measured productivity values. Recent satellite-derived estimates of surface ocean productivity (for example, ref. 30) tend to have higher absolute values than those in the previous literature. Satellite mapping is still developing, but this could mean that ultimately the values to which we have calibrated will need to be adjusted upwards. However, the trends and relative changes in productivity we used follow CZCS imagery and are compatible with satellite observations. Within the context of the calibration used here, estimation error is about 10% of the calculated value²⁸.

An average of 346 individuals was counted for each downcore sample and the abundances of 33 taxa were used to make palaeoestimations. Counts for the four cores were made independently by three researchers (P. L., M. Fariduddin, Q. Hui) trained in the taxonomic system used to calibrate the transfer function. Tests for no-analogue conditions, following ref. 12, show that the downcore assemblages can be modelled in terms of surface calibration data.

Received 10 September 1999; accepted 25 May 2000.

- Barber, R. & Chavez, R. Regulation of primary productivity rate in the equatorial Pacific. *Limnol. Oceanogr.* **36**, 1803–1815 (1991).
- Tans, P., Fung, I. & Takahashi, T. Observational constraints on the global atmosphere CO₂ budget. *Science* **247**, 1431–1438 (1990).
- Murray, J., Barber, R., Roman, M., Bacon, M. & Feely, R. Physical and biological controls on the carbon cycling in the equatorial Pacific. *Science* **266**, 58–65 (1994).
- Toggweiler, J., Dixon, D. & Broecker, W. The Peru upwelling and the ventilation of the South Pacific thermocline. *J. Geophys. Res.* **96**, 20467–20497 (1991).
- Pisias, N., Mix, A. & Zahn, R. Non-linear response in the global climate system: evidence from benthic oxygen isotope record in core RC13-110. *Paleoceanography* **5**, 147–160 (1990).
- Mix, A. & Shackleton, N. Benthic foraminiferal stable isotope stratigraphy of site 846: 0–1.8 M.A. *Proc. ODP Sci. Res.* **138**, 839–854 (1995).
- Ninkovitch, D. & Shackleton, N. Distribution, stratigraphic position and age of ash layer 'L' in the Panama Basin region. *Earth Planet. Sci. Lett.* **27**, 20–34 (1975).
- Loubere, P. & Fariduddin, M. Quantitative estimation of global patterns of surface ocean biological productivity and its seasonal variation on timescales from centuries to millennia. *Glob. Biogeochem. Cycles* **13**, 115–133 (1999).
- Lampitt, R. & Antia, A. Particle flux in the deep seas: regional characteristics and temporal variability. *Deep Sea Res.* **44**, 1377–1403 (1997).
- Loubere, P. The surface ocean productivity and bottom water oxygen signals in deep water benthic foraminiferal assemblages. *Mar. Micropaleontol.* **28**, 247–262 (1996).
- Fariduddin, M. & Loubere, P. The surface ocean productivity response of deeper water benthic foraminifera in the Atlantic Ocean. *Mar. Micropaleontol.* **32**, 289–310 (1997).
- Loubere, P. A multiproxy reconstruction of biological productivity and oceanography in the eastern equatorial Pacific for the past 30,000 years. *Mar. Micropaleontol.* **37**, 173–198 (1999).
- Lyle, M. *et al.* The record of Late Pleistocene biogenic sedimentation in the eastern tropical Pacific Ocean. *Paleoceanography* **3**, 39–59 (1988).
- Pisias, N. & Mix, A. Spatial and temporal oceanographic variability of the eastern equatorial Pacific during the late Pleistocene: evidence from radiolarian microfossils. *Paleoceanography* **12**, 381–394 (1997).
- Martin, J. Glacial-interglacial CO₂ change: the iron hypothesis. *Paleoceanography* **5**, 1–13 (1990).
- Kutzbach, J. *et al.* Climate and biome simulations for the past 21,000 years. *Quat. Sci. Rev.* **17**, 473–506 (1998).
- Rea, D. The paleoclimatic record provided by aeolian deposition in the deep sea: the geologic history of wind. *Rev. Geophys.* **32**, 159–195 (1994).
- Mix, A. & Morey, A. in *The South Atlantic: Present and Past Circulation* (eds Wefer, G. *et al.*) 503–525 (Springer, Berlin, 1996).
- Liu, Z. *et al.* Dynamic and observational constraints on the tropical Pacific Ocean at the last glacial maximum. *Geophys. Res. Lett.* **27**, 105–108 (2000).
- Francois, R. *et al.* Contribution of Southern Ocean surface water stratification to low atmospheric CO₂ concentrations during the last glacial period. *Nature* **389**, 929–935 (1997).
- Toggweiler, J. & Carson, S. in *Upwelling in the Ocean: Modern Processes and Ancient Records* (eds Summerhayes, C. *et al.*) 337–359 (Wiley and Sons, New York, 1995).
- Barber, R. *et al.* Primary productivity and its regulation in the equatorial Pacific during and following the 1991–1992 El Niño. *Deep Sea Res.* **43**, 933–969 (1996).
- Fitzwater, S., Coale, K., Gordon, R., Johnson, K. & Ondrusek, M. Iron deficiency and phytoplankton growth in the equatorial Pacific. *Deep Sea Res.* **43**, 995–1015 (1996).
- Farrell, J., Pedersen, T., Calvert, S. & Nielsen, B. Glacial-interglacial changes in nutrient utilization in the equatorial Pacific Ocean. *Nature* **377**, 514–517 (1995).
- Dugdale, R. & Wilkerson, F. Silicate regulation of new production in the equatorial Pacific upwelling. *Nature* **391**, 270–273 (1998).

- Berger, W. & Lange, C. Silica depletion in the thermocline of the glacial North Pacific: corollaries and implications. *Deep Sea Res.* **45**, 1885–1904 (1998).
- Loubere, P. Quantitative estimation of surface ocean productivity and bottom water oxygen concentration using benthic foraminifera. *Paleoceanography* **9**, 723–737 (1994).
- Loubere, P. The impact of seasonality on the benthos as reflected in the assemblages of deep-sea foraminifera. *Deep Sea Res.* **45**, 409–432 (1998).
- Berger, W. in *Productivity of the Ocean: Present and Past* (eds Berger, W., Smetacek, V. & Wefer, G.) 429–455 (Wiley, New York, 1989).
- Antoine, D., Jean-Michel, A. & Morel, A. Oceanic primary production. 2. Estimation at global scale from satellite (coastal zone color scanner) chlorophyll. *Glob. Biogeochem. Cycles* **10**, 57–69 (1996).

Acknowledgements

We thank M. Fariduddin and Q. Hui for helping to generate the benthic count data. This work was supported by the NSF.

Correspondence and requests for materials should be addressed to P.L. (e-mail: paul@geol.niu.edu).

Annual modulation of triggered seismicity following the 1992 Landers earthquake in California

Stephen S. Gao^{*†}, Paul G. Silver^{*}, Alan T. Linde^{*} & I. Selwyn Sacks^{*}

^{*} Department of Terrestrial Magnetism, Carnegie Institution of Washington, 5241 Broad Branch Road, NW, Washington DC 20015, USA

The mechanism responsible for the triggering of earthquakes remains one of the least-understood aspects of the earthquake process. The magnitude-7.3 Landers, California earthquake of 28 June 1992 was followed for several weeks by triggered seismic activity over a large area, encompassing much of the western United States¹. Here we show that this triggered seismicity marked the beginning of a five-year trend, consisting of an elevated microearthquake rate that was modulated by an annual cycle, decaying with time. The annual cycle is mainly associated with several hydrothermal or volcanic regions where short-term triggering was also observed. These data indicate that the Landers earthquake produced long-term physical changes in these areas, and that an environmental source of stress—plausibly barometric pressure—might be responsible for the annual variation.

For this study, we used the CNSS (Council of the National Seismic System) earthquake catalogue for the period 1 January 1988 to 1 January 1997, with the aftershocks removed by declustering² (Fig. 1). Figure 2a shows the resulting time series of seismicity rate r (number of events per day). Before the Landers earthquake, r is relatively stable, averaging 39.7 events per day. Just afterwards, there is a sharp peak in r due to remotely triggered seismicity from the Landers event, which has been extensively studied^{1,3–10}. This is followed by a remarkable annual periodicity that appears to modulate the slow decay of a post-Landers boost in seismicity (Fig. 2). The mean peak-to-peak amplitude of the annual cycle, averaged over the 4-year period 1 January 1993 to 1 January 1997 is about 10 events per day (Fig. 3). If modelled as a decaying sinusoid of the form

$$r(t) = a_1 + e^{-a_2 t} [a_3 + a_4 \sin(2\pi t - a_5)] \quad (1)$$

where t is time in years since the beginning of 1993, a decay constant of about 2 years is obtained (Fig. 2b). We tested the significance of the annual cycle by fitting the post-Landers data using equation (1), and found that a_4 is significantly different from zero at the 95% level

[†] Present address: Department of Geology, Kansas State University, Manhattan, Kansas 66506, USA.

for cut-off magnitudes of 1.5 and lower (Fig. 2d). The excess seismicity and its decay could not be the result of incomplete removal of aftershocks of major earthquakes such as the Landers, Big Bear, Joshua Tree and Little Skull Mountain earthquakes, because these features still exist when these aftershock zones are excluded.

The annual periodicity in seismicity is spatially localized by calculating the numbers of events N_{e1} and N_{e2} that occurred in the first half (months 1–6) and second half (months 7–12) of the years, respectively, for both the pre- and post-Landers periods. Based on Fig. 3, the difference $\Delta N_e = N_{e2} - N_{e1}$ should be highest for the post-Landers period in areas that possess a large annual cycle with a peak in the autumn. In Fig. 4, we have mapped ΔN_e for the post-Landers period. It is clear that the largest contribution comes from three areas: Long Valley, Coso and The Geysers geothermal field. These are all hydrothermal/volcanic areas and were subject to the most intense short-term triggering from Landers. This suggests a close link between the short-term triggering and the long-term seismicity. All together these three areas contribute about 2/3 of the total amplitude of the annual cycle (Fig. 3).

The variations in seismicity are detected only below magnitude 1.5 (Fig. 2d). Because for most areas, the completeness magnitude M_c (the magnitude above which no events are missed) is thought to be in the range 1.1–2.5, the catalogue is incomplete for most of the areas. It is thus necessary to determine whether the observed annual cycle could be an artefact of temporal variations in network sensitivity after the Landers earthquake compared to the preceding period. This would require the existence of an annual cycle in network sensitivity with the maximum in the autumn and mini-

mum in the spring. A particularly sensitive measurement of network sensitivity is the seasonal variation in the number of stations that detect events at fixed magnitude. Figure 5 suggests that in the pre-Landers period, for the entire area, there are (on average) one or two more stations per event in the second half of the years than in the first half. If this made a noticeable difference in observed seismicity rate, we would expect to see an annual cycle during the pre-Landers period with peaks in the autumn. This is not the case (Figs 2b, 3). For the post-Landers period, there are several more stations per event in the first half of the years than in the second half. Again, if this made a noticeable difference in seismicity rate, we would expect to see an annual cycle with more events in the first half of the years, which is the opposite of what we have observed.

To study local variations of network sensitivity, we first divided the area into equal-sized blocks; then, for each block, we computed the number of stations per event N_{s1} , N_{s2} for the first half and second half of the years, respectively, and took the difference $\Delta N_s = N_{s2} - N_{s1}$. The spatial distribution of the resulting ΔN_s suggests that there is a great amount of spatial variability in ΔN_s . If the temporal variations in seismicity were due to variations in network sensitivity, the maps of ΔN_s and ΔN_e (Fig. 4) should be very similar. In fact, the two patterns are essentially uncorrelated. One area where ΔN_s is positive for small-magnitude events (below $M = 1.5$) is in the Long Valley area. In the vicinity of the caldera, ΔN_s varies between 1 and 2 (out of 10–20 total stations), as might be expected due to heavy snow in the winter. A detailed analysis of the seismicity of this particular zone, however, shows that the post-Landers seismicity that shows an annual cycle is primarily in the region directly south of the caldera, where ΔN_s is effectively zero. In

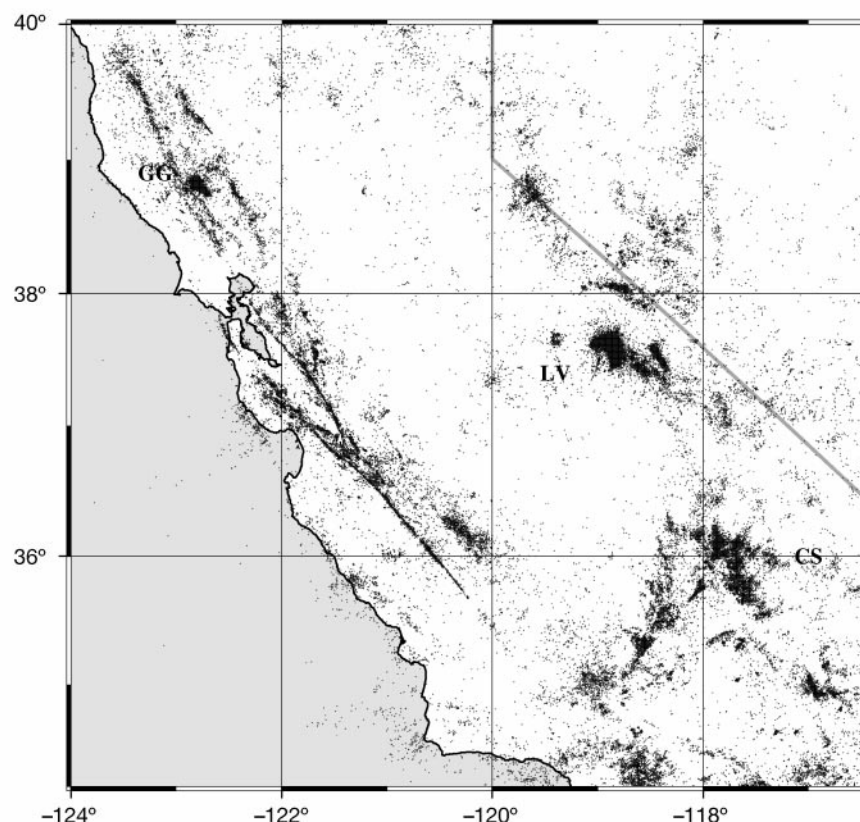


Figure 1 Map showing declustered seismicity in the western United States for the period 1 January 1988 to 1 January 1997. The total number of events in the area is about 141,000. We use a cluster-link procedure for the declustering². The dimensionless scaling factor $Q = 10$ is used in the study, although our tests show that the presence of

the annual cycle is insensitive to the particular value of Q used (values between 2 and 20 were used). GG, The Geysers geothermal area; LV, Long Valley; and CS, Coso. The 1992 Landers mainshock is located at the lower-right corner of the plot.

addition, considering together this zone plus the other areas with a large annual cycle, the average ΔN_s below $M = 1.5$ is very small (a few tenths of a station) and the values for the pre- and post-Landers periods are statistically indistinguishable (Fig. 5). These regions taken together nevertheless continue to reveal a clear annual cycle. Finally, this seismicity is dominated by the occurrence of swarms. It is thus highly unlikely that subtle temporal variations in network sensitivity could be the cause. The possibility that the annual cycle was the result of seasonal variation in water pumping/injection after June 1992 can also be ruled out (E. Johnson, M. Khan and S. Hodgson, personal communications).

We find that most of the swarms in this area and in the other hydrothermal/volcanic areas occurred more often in the autumn. It is this preference in the occurrence time of swarms that gives rise to much of the observed annual cycle (Fig. 2). Based on the above analysis, we conclude that variable network sensitivity is an unlikely explanation for the observed characteristics of seismicity. We propose that a real, post-Landers increase in small-event seismicity, modulated by an annual cycle, is the most likely explanation for the observed annual cycle in earthquake activity.

The data presented here reveal the existence of a long-term component of the previously observed Lander-triggered seismicity, which is modulated by an annual cycle. The excess seismicity implies a long-term change in the conditions for failure at great distance from the Landers earthquake. The predicted static stresses from the Landers earthquake are small at large distances, of the

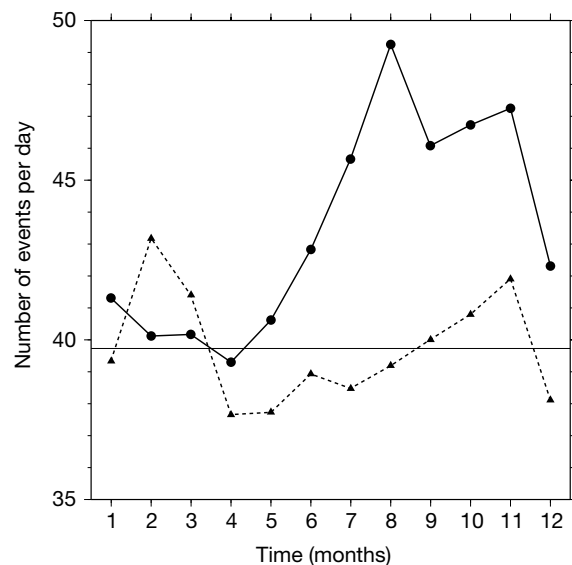


Figure 3 Average amplitude of annual modulation. Solid curve, monthly averages of seismicity-rate for the 4-year period from 1 January 1993 to 1 January 1997; dashed curve, averages for 4-year period from 1 January 1988 to January 1992. Post-Landers minimum is approximately equal to pre-Landers mean value (horizontal line).

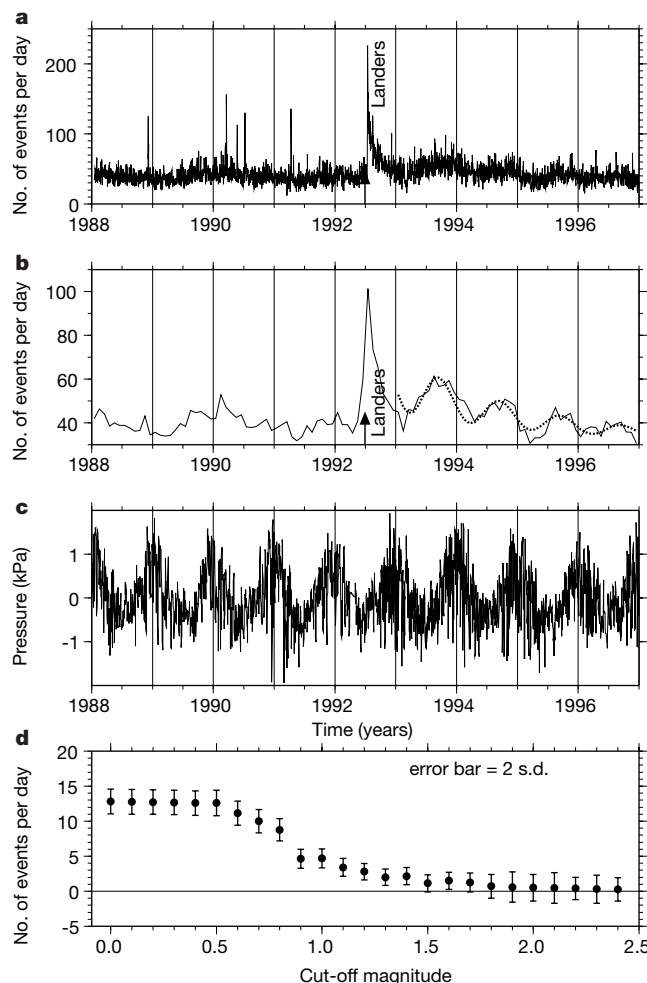


Figure 2 Time series of seismicity-rate and amplitude of the annual cycle. **a**, Number of events per day. In **b**, the solid line shows the 30-day running mean of the number of events per day, and the dotted line is the fit to equation (1). **c**, Daily mean barometric

pressure recorded at a site near Long Valley. **d**, Amplitude of the annual cycle as quantified by a_4 in equation (1), calculated using different cut-off magnitudes.

order of 1 kPa (10 mbar), whereas the triggering of aftershocks appear to require a Coulomb stress that is an order of magnitude greater. The dynamic stresses (due to propagating seismic waves) from the event, however, were capable of generating 10-kPa stress changes, and this mechanism has been invoked to explain the short-term triggering^{1,3,9}. If this same mechanism is responsible for the long-term triggering, it implies that dynamic stresses can produce semi-permanent changes at large distances from a seismic event.

The existence of the annual cycle requires a process that is capable of modulating the Coulomb stress for small seismic events over a large area. The most plausible sources of stress with a large annual cycle are barometric pressure and precipitation (tidal stresses at one-year period are about two orders of magnitude lower). We regard barometric pressure as the more likely of the two. This is primarily because the levels of precipitation vary dramatically throughout the regions (from large snowfall to desert environments), whereas barometric pressure fluctuations are much more coherent. Comparison of seismicity with pressure (Fig. 2) reveals that pressure lows precede seismicity highs by 90 days. Barometric pressure applies a time-varying surface load with dominant annual period and amplitude of about 2 kPa. For vertical faults, a pressure low will produce a decrease in normal stress and a corresponding increase in the effective shear stress that should produce a high in seismicity. The presence of a phase lag could be explained by the effect of fluid pore pressure on Coulomb stress. This effect is time dependent, owing to diffusion to the water table¹¹. A hydrologic diffusivity of $1 \text{ m}^2 \text{ s}^{-1}$ —which is a high but not unreasonable value for fractured volcanic rock that characterizes the triggered regions—could produce the observed lag.

The above analysis suggests that the triggered seismicity from the Landers earthquake possesses a long-term, multi-year component. The annual periodicity is primarily observed in hydrothermal/volcanic areas. Barometric-pressure-induced modulation of seismicity may explain these data. Such a stress variation (about 2 kPa) is an order of magnitude smaller than the minimum static stress levels previously considered to trigger aftershocks^{12,13}. □

Received 4 October 1999; accepted 8 June 2000.

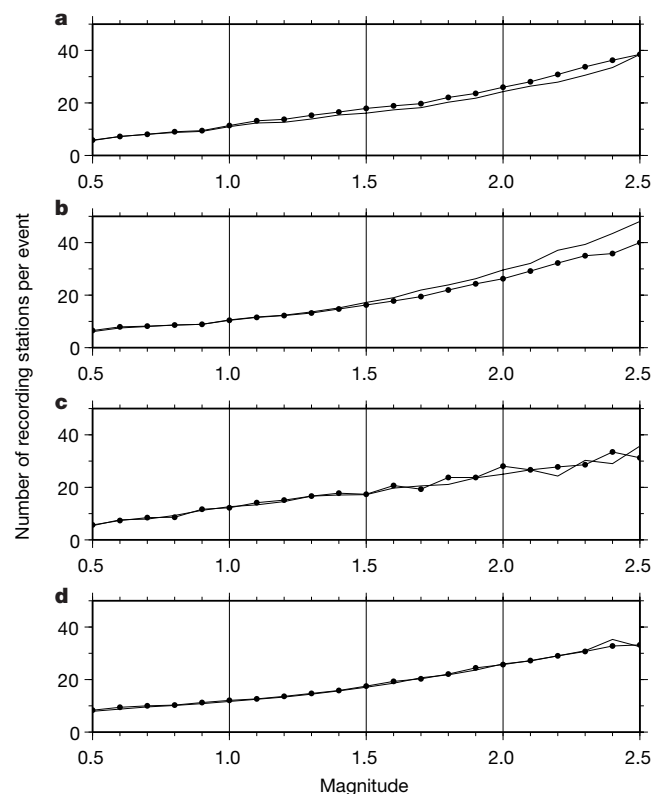


Figure 5 Number of recording stations plotted as a function of earthquake magnitude for the study area. Panels **a** and **b** show results for the entire study area. **a**, pre-Landers; **b**, post-Landers. Lines through dots are for the second half of the years, and lines without dots are for the first half of the years. Panels **c** and **d** are as **a** and **b** but for volcanic/hydrothermal regions exhibiting large annual cycles: these are the three red regions in Fig. 4, excluding northern Long Valley (see text).

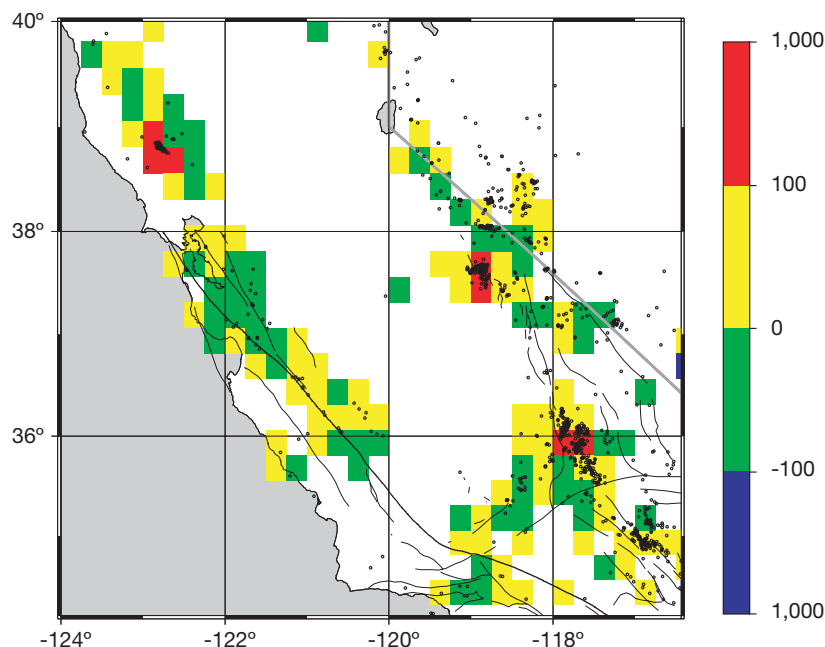


Figure 4 Map showing spatial distribution of excess seismicity in the second half of the years relative to the first half of the years for the post-Landers period. The values shown on the colour scale are the difference in the total number of events between the second

half and first half of the years for the post-Landers period. Dots are events that occurred within 10 days after Landers. We note high excess seismicity in areas where short-term triggering areas was produced by the Landers earthquake.

- Hill, D. P. *et al.* Seismicity remotely triggered by the magnitude 7.3 Landers, California, earthquake. *Science* **260**, 1617–1623 (1993).
- Reasenber, P. A. Second-order moment of central California seismicity 1969–1982. *J. Geophys. Res.* **90**, 5479–5495 (1985).
- Gomberg, J. & Davis, S. Stress/strain changes and triggered seismicity at The Geysers, California. *J. Geophys. Res.* **101**, 733–749 (1996).
- Hill, D. P., Johnston, M. J. S., Langbein, J. O. & Bilham, R. Response of Long Valley caldera to the $M_w = 7.3$ Landers, California, earthquake. *J. Geophys. Res.* **100**, 12985–13005 (1995).
- Anderson, J. G. *et al.* Seismicity in the western Great Basin apparently triggered by the Landers, California, Earthquake, 28 June 1992. *Bull. Seismol. Soc. Am.* **84**, 863–891 (1994).
- Gomberg, J. & Bodin, P. Triggering of the $M_s = 5.4$ Little Skull Mountain, Nevada, earthquake with dynamic strains. *Bull. Seismol. Soc. Am.* **84**, 844–853 (1994).
- Linde, A. T. Increased pressure from rising bubbles as a mechanism for remotely triggered seismicity. *Nature* **371**, 408–410 (1994).
- King, G. C. P., Stein, R. S. & Lin, J. Static stress changes and the triggering of earthquakes. *Bull. Seismol. Soc. Am.* **84**, 935–953 (1994).
- Gomberg, J., Beeler, N. M., Blanpied, M. L. & Bodin, P. Earthquake triggering by transient and static deformations. *J. Geophys. Res.* **103**, 24411–24426 (1998).
- Sturtevant, B., Kanamori, H. & Brodsky, E. E. Seismic triggering by rectified diffusion in geothermal systems. *J. Geophys. Res.* **101**, 25269–25282 (1996).
- Roeloffs, E. in *Advances in Geophysics* Vol. 37 (ed. Dmowska, R.) 135–195 (Academic, San Diego, California, 1996).
- Stein, R. S. & Lisowski, M. The 1979 Homestead Valley earthquake sequence, California; control of aftershocks and postseismic deformation. *J. Geophys. Res.* **88**, 6477–6490 (1983).
- Reasenber, P. A. & Simpson, R. W. Response of regional seismicity to the static stress change produced by the Loma Prieta earthquake. *Science* **255**, 1686–1690 (1992).

Acknowledgements

The earthquake catalogue was obtained from the Northern California Data Center maintained by UC Berkeley and the USGS. The code for declustering catalogues was written by P. Reasenber. We thank the National Climatic Data Center and E. Roeloffs and E. Quilty for providing precipitation data; E. Roeloffs, D. Oppenheimer, S. Solomon, L. Knopoff, E. Press and H. Liu for discussions; and J. Gomberg for comments that significantly improved the manuscript.

Correspondence and requests for materials should be addressed to S.S.G. (e-mail: sgao@ksu.edu).

Ocean circulation off east Antarctica affects ecosystem structure and sea-ice extent

Stephen Nicol*, Tim Pauly*, Nathan L. Bindoff†, Simon Wright*, Deborah Thiele‡, Graham W. Hosie*, Peter G. Strutton§|| & Eric Woehler*

* Australian Antarctic Division, Department of the Environment and Heritage, Channel Highway, Kingston, Tasmania 7050, Australia

† Antarctic Co-operative Research Centre, University of Tasmania, GPO Box 252C-80 Hobart, Tasmania 7001, Australia

‡ School of Ecology and Environment, PO Box 423, Warrnambool, Victoria 3280, Australia

§ School of Biology, Flinders University of South Australia, Adelaide, South Australia 5001, Australia

Sea ice and oceanic boundaries have a dominant effect in structuring Antarctic marine ecosystems. Satellite imagery and historical data have identified the southern boundary of the Antarctic Circumpolar Current¹ as a site of enhanced biological productivity². Meso-scale surveys off the Antarctic peninsula have related the abundances of Antarctic krill (*Euphausia superba*) and salps (*Salpa thompsoni*) to inter-annual variations in sea-ice extent³. Here we have examined the ecosystem structure and oceanography spanning 3,500 km of the east Antarctic coast-

line, linking the scales of local surveys and global observations. Between 80° and 150° E there is a threefold variation in the extent of annual sea-ice cover, enabling us to examine the regional effects of sea ice and ocean circulation on biological productivity. Phytoplankton, primary productivity, Antarctic krill, whales and seabirds were concentrated where winter sea-ice extent is maximal, whereas salps were located where the sea-ice extent is minimal. We found enhanced biological activity south of the southern boundary of the Antarctic Circumpolar Current rather than in association with it². We propose that along this coastline ocean circulation determines both the sea-ice conditions and the level of biological productivity at all trophic levels.

The BROKE (baseline research on oceanography, krill and the environment) survey was designed to obtain a quasi-synoptic description of the oceanography and ecosystems in an area of 873,000 km² off the coast of east Antarctica during the austral summer of 1996 (ref. 4; Fig. 1). Underway measurements included near-surface ocean currents, sea surface temperature, multi-frequency acoustics, surface water fluorescence and whale and seabird observations. On eight transects, detailed oceanographic measurements, primary productivity measurements, phytoplankton sampling and oblique 0–200-m zooplankton net tows were carried out at pre-determined stations.

Surface circulation between 80° and 150° E was characterized by a narrow westward slope current just north of the shelf break following bathymetric contours (Fig. 1). Flow was westward along the shelf, northwards near 80° E, with an eastward return forming a gyre closed at 115–120° E. Along the southern boundary of the Antarctic Circumpolar Current (SB-ACC), the flow was highly variable but primarily eastwards. Near 130° E, north of the SB-ACC, the flow was more strongly eastwards. This circulation pattern confirmed measurements from iceberg tracks⁵ and sea-ice drift⁶. Sea surface temperatures were cooler in the west where the cool water flowed around the gyre and penetrated to lower latitudes (Fig. 1). The survey crossed the SB-ACC¹ on five transects, and with one exception the position of the SB-ACC was close to previous measurements, being further offshore in the west and closer to the coast in the east (Fig. 1).

Significantly more of the acoustically detected Antarctic krill biomass (63% or 3.0 million tonnes) was encountered in the west of the survey area (80–115° E; Table 1). In the east, Antarctic krill were confined to the narrow coastal band of cold (less than –1.6°C) Antarctic Surface Water (AASW; Fig. 2). Antarctic krill biomass in the west was distributed throughout the latitudinally extensive cool surface waters south of the SB-ACC. Not many studies have examined Antarctic krill distribution in this region. The *Discovery*

Table 1 Biological measurements in the west (80–115° E) and the east (115–150° E) of the BROKE survey area

	Mean west (n)	Mean east (n)	F	P	d.f.
Krill acoustics*	34.39 (1076)	22.48 (927)	6.142	0.013	1,2001
Salps density (RMT net)†	0.72 (29)	38.8 (37)	65.247	< 0.0001	1,64
Primary productivity‡	613.4 (14)	395.36 (14)	11.28	0.002	1,26
Chlorophyll-a§	68.36 (29)	27.93 (31)	13.624	0.0005	1,58
Birds	70.85 (187)	30.73 (127)	5.561	0.019	1,312
Whales¶	0.066 (9)	0.003 (9)	6.181	0.024	1,16

Comparisons made by analysis of variance (ANOVA).

* Mean acoustic sa (m² per nautical square mile) for bins of 10 nautical miles.

† Density in individuals per 1,000 m³ (log₁₀ (x+1) transformed before analysis).

‡ Column productivity in mg C m⁻² d⁻¹ from CTD casts.

§ Vertically integrated chlorophyll a for 0–150 m in mg Chl-a m⁻² from CTD casts.

|| Number of birds observed in 10-minute intervals on the hour, standardized for effort.

¶ Mean number of cetaceans observed per transect, standardized for effort.

|| Present address: Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, California 95039, USA.

Reports describe Antarctic krill between 85° and 100° E as being “plentiful”, but between 120° and 150° E as being confined close to the coast in the East Wind Drift, which “here forms a narrower belt near the coast than anywhere else”⁷.

Seabirds and baleen whales, chief predators on Antarctic krill, shared a distribution pattern that reflected that of their prey (Fig. 2, Table 1). Humpback whales were the most frequently sighted cetacean accounting for 26.7% of all sightings. All humpback whales (77 sightings, 150 individuals) and 85.1% of all cetacean sightings occurred in the west and in the east, cetaceans were most abundant in the cooler coastal waters (Fig. 2). This pattern reflects the distribution of blue and humpback whales from historical catch records^{2,8}.

Chlorophyll-*a* concentrations and primary productivity were also highest in a broad band in the west where the SB-ACC was further offshore (Fig. 3, Table 1). In the east, chlorophyll-*a* and productivity were confined to a narrow coastal band and here the SB-ACC was closest to the coast. The observed pattern of

phytoplankton productivity and its distribution matches the long-term average chlorophyll abundance for the region derived from satellite imagery⁹, suggesting that the east–west differences are not merely a seasonal effect.

Biological activity off east Antarctica was distributed between the SB-ACC and the ice edge, and was not significantly associated with the SB-ACC itself as has been previously suggested². Primary productivity at stations closest to the SB-ACC (mean 283 mg C m⁻² d⁻¹, *n* = 8), was not significantly different from all other stations (mean 313 mg C m⁻² d⁻¹, *n* = 20, *t* = 0.7309, critical value 2.056). Fluorescence distribution peaked at a surface water temperature of -0.54 ± 0.90 °C (mean $\pm$ s.d.) far south of the SB-ACC. The null hypothesis that fluorescence data are uniformly distributed over the range of temperature observations was rejected ($\chi^2 = 444.5$, $\alpha = 0.01$, 29 degrees of freedom (d.f.)). Krill distribution peaked at a surface water temperature of -0.01 ± 1.00 °C, far south of the SB-ACC (Fig. 1). The null hypothesis that the acoustic back-scatter data are uniformly distributed over the range of temperature observations was rejected ($\chi^2 = 767.2$, $\alpha = 0.01$, 19 d.f.). Whale distribution was negatively correlated with vertically integrated temperature (pairwise Pearson correlation matrix, *r* = -0.82), and whales were mainly found in the cooler waters south of the SB-ACC.

The distribution of primary production off east Antarctica was directly reflected in the large-scale distributions of Antarctic krill, and of their predators—baleen whales and seabirds. Smaller-scale

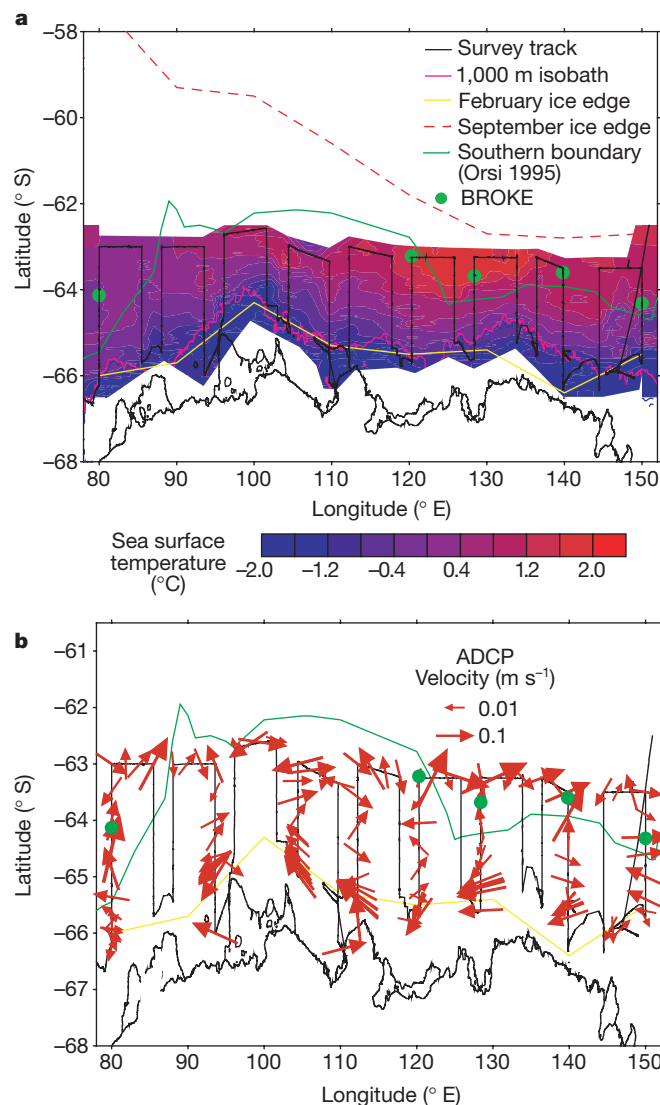


Figure 1 Oceanographic features off east Antarctica. **a**, The BROKE survey track showing the reported position of the southern boundary of the Antarctic Circumpolar Current¹, and the position derived from BROKE oceanographic measurements, sea surface temperature from underway measurements and the average sea-ice extent in February (minimum) and September (maximum) derived from satellite data³⁰. **b**, Surface circulation over the BROKE survey area. Current vectors are smoothed velocities from the acoustic Doppler current profiler (ADCP) interpolated to all the oceanographic stations at the 52.6-m depth level.

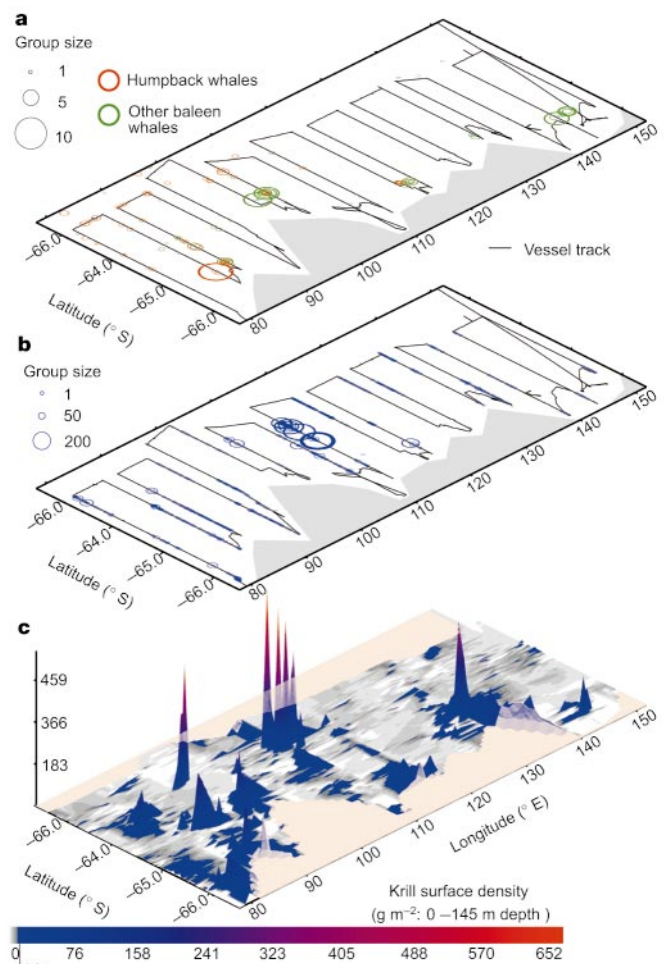


Figure 2 Distribution and abundance of whales, seabirds and Antarctic krill off east Antarctica, austral summer 1996. **a**, Whales. **b**, Seabirds. **c**, Antarctic krill (*Euphausia superba*) in grams per square metre. BROKE survey track is indicated.

relationships between Antarctic krill and phytoplankton distributions¹⁰, and Antarctic krill and krill predators¹¹, have been examined previously but large-scale patterns have been limited to interpretations of historical datasets². To our knowledge, our results provide the first direct observations from the Antarctic of large-scale spatial linkages between the oceanography and primary productivity, and the distribution and abundance of Antarctic krill and their predators.

Off the Antarctic peninsula, the pelagic zone in summer appears to be dominated by Antarctic krill after winters with above average sea-ice conditions, or by salps after winters with below average sea-ice conditions³. Off east Antarctica, there was a geographic expression of this temporal relationship: salps were significantly more abundant in the region with the least annual sea-ice cover (115–150° E), whereas Antarctic krill were most abundant in the region with the greatest annual sea-ice cover (80–115° E) (Figs 1, 2 and 3; Table 1). Salps were virtually absent from the productive waters in the west of the survey area and were mainly found in the east of the survey area where warmer, low chlorophyll, oceanic waters approached the continental shelf (Fig. 3). Other studies^{12–14} have reported a similar lack of spatial overlap between salps and Antarctic krill in the Southern Ocean. Antarctic krill appear to require high, patchy, chlorophyll concentrations¹⁵, whereas salps generally inhabit uniformly low chlorophyll oceanic waters¹⁶ implying habitat separation rather than competition³¹.

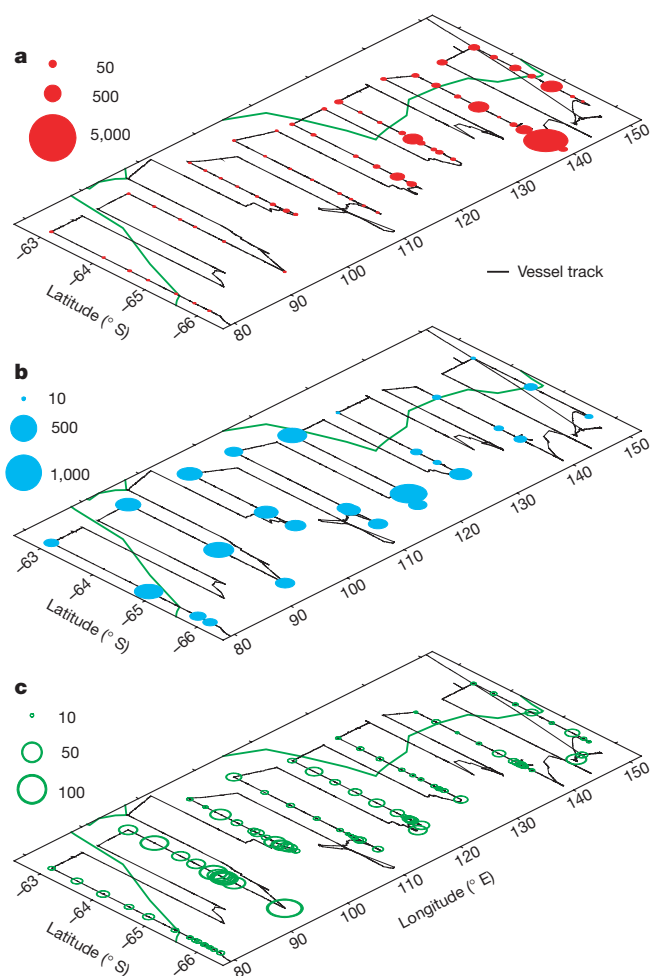


Figure 3 Distribution and abundance of salps, primary productivity and chlorophyll-*a* off east Antarctica, austral summer 1996. **a**, Salp (*Salpa thompsoni*) density (individuals per 1,000 m³). **b**, Gross production in mg C per square metre per day. **c**, Chlorophyll-*a* stocks in milligrams per square metre. The BROKE survey track and the reported position of the SB-ACC¹ are indicated.

On the basis of our observations, we propose a conceptual model that links oceanography, sea ice, and the relative abundances of Antarctic krill and salps off east Antarctica (Fig. 4). Ocean circulation is seen to drive the annual sea-ice extent, the position of oceanic boundaries and biological productivity. Consequently, there is a positive relationship between the distance offshore of the SB-ACC and the maximum extent of winter sea ice (ref. 2; Fig. 1). Where the coastal current develops a northward component it becomes wider, the SB-ACC is found further offshore and there is greater seasonal sea-ice cover. There is also a wider band of biological activity, containing abundant Antarctic krill and krill predators, and high primary productivity. Salps inhabit the warmer, low chlorophyll, oceanic waters of the ACC and are therefore found close to the continent only in the area where the coastal current is narrowest. Here the SB-ACC comes close to the coast, and consequently, seasonal sea ice is minimal.

The Antarctic marine ecosystem is affected by variability in the sea-ice and oceanographic conditions³; this has been linked to both long-term trends¹⁷ and shorter-term cycles in sea-ice extent and sea surface temperature¹⁸. Temporal relationships between oceanography, sea ice and biology observed in other regions of the Antarctic may be explained by using our conceptual model for east Antarctica. If the position of the SB-ACC can vary significantly between years, as has been documented in the South Atlantic^{19,31}, this could account for inter-annual biological and physical variability observed in other areas. Our conceptual model would predict that in a given region, in years where the SB-ACC is located further offshore, cooler coastal waters would be more extensive, the area would experience above average winter sea-ice conditions, Antarctic krill populations would be more extensive and salps would be found further offshore. Furthermore, when the SB-ACC is located further inshore, the cool coastal current would be narrower, winter sea ice would be less extensive, Antarctic krill would be scarce, or more coastsly constrained, and salps more abundant closer to shore.

Understanding the relationships between the dynamics of the Antarctic marine ecosystem and changes in the physical environment is critical for predicting the regional effects of climate change¹⁷. Antarctic krill is also subject to the region's largest fishery and development of sustainable management strategies for this species requires information on inter-annual variability of the stocks and of the factors that drive such changes²⁰. Large surveys, such as BROKE, which encompass principal regional differences in the physical environment, can be critical in assessing the generality of hypotheses generated from regular small-scale surveys and in producing ground-truth data to validate global-scale satellite observations.

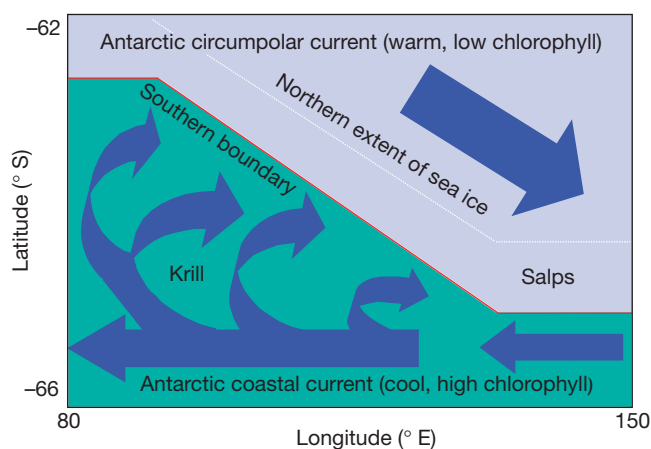


Figure 4 Conceptual model relating the circulation pattern derived from ocean current measurements, the sea-ice extent and the distribution and abundance of chlorophyll, salps and Antarctic krill off east Antarctica.

Methods

BROKE was conducted between 29 January and 22 March 1996 on the RSV *Aurora Australis* and consisted of 18 transect lines extending, on average, 239 km from the ice edge between 80° and 150° E.

Oceanography

An average of 13 CTD (conductivity, temperature and depth) casts per transect were carried out on eight north–south transects measuring temperature, salinity and other variables. Station spacing over the continental slope was as little as 1 nautical mile. Underway measurements included ocean currents using an acoustic Doppler current profiler and sea surface temperature.

Acoustics

Acoustic data were collected at 8 knots using Simrad EK500 echosounders with hull-mounted transducers (200, 120, 38 and 12 kHz). Acoustic mean volume backscatter data (Sv) for 120 kHz were continuously logged from single pings at 700 vertical samples over a 250 m depth layer below the transducer. All on-transect data (except CTD and trawl stations) were used to estimate Antarctic krill abundance. Sv data were integrated over 5-m layers between 5 and 245 m depth below the transducer, averaged over intervals of 0.1 nautical miles. Background noise was subtracted from the signal. The transect mean Sv and weights were used to calculate mean Sv and variance for the survey²¹. Mean Antarctic krill target strength and weight were calculated from individual Antarctic krill²² and 113 net trawls were directed to identify acoustically detected targets.

Cetacean sightings

We used a double-platform line transect passing mode survey method²³. Sighting shifts of 30 min were conducted throughout daylight hours, weather dependent. A modified single platform survey methodology operated when the ship stopped or slowed and in sea states ≥ 4 . The total searching effort was 112.45 hours west of 120° E, and 99.9 hours in the east of and including 120° E. The percentage of searching effort in Beaufort sea states ≤ 3 was 71% in the west and 67.2% in the east. Thus, search effort and sightings probability with respect to effort and sea state are comparable between the two areas. In total 288 sightings were made, consisting of 574 individuals.

Bird observations

Observations of seabird numbers within a 300-m forward quadrant of the ship were recorded continuously during daylight hours while the ship was underway when visibility was beyond 500 m (ref. 24).

Pigments

Phytoplankton were collected by filtering 200–2,000 ml of seawater from Niskin bottles at 15-m depth intervals from 0 to 200 m at CTD stations through Whatman GF/F filters in subdued light. Pigments were extracted by sonication in 1.0–1.5 ml 100% methanol, using 130 ng apo-8'-carotenol (Fluka) as an internal standard and analysed by high performance liquid chromatography (HPLC) and were detected using Waters 996 photodiode array and Hitachi FT1000 fluorescence detectors. Pigments were identified by their ultraviolet–visible spectra, and by comparison of retention times with a standard pigment mixture from SCOR reference algae²⁵.

Primary production

We carried out 28 small-bottle ¹⁴C-productivity versus irradiance (PvsI) incubations²⁶. Usually 4 depths per station between the surface and 75 m were sampled; continuous chlorophyll profiles were derived from linear interpolation of the discrete HPLC chlorophyll profiles. After standardizing the individual data points on each PvsI curve to chlorophyll-*a*, we used a nonlinear parameter estimation routine to fit PvsI models^{27,28}. The gross daily depth-integrated column production was then calculated²⁸ using climatological surface irradiance²⁹. Depth-integrated production was calculated from the surface to the bottom of the shallow mixed layer (*Z_m*), and from the bottom of the mixed layer to the maximum depth sampled. Primary productivity integrated to 125 m is shown in Fig. 3 and used for the statistical analysis in Table 1.

Net sampling

Sixty-three downward oblique hauls (0–200 m) with a rectangular mid-water trawl (8-m² mouth area, 4.5-mm mesh) were conducted on eight transects, and on four additional sites between transects to determine zooplankton community structure.

Analyses

Productivity stations were divided into those closest to the SB-ACC ($< 0.25^\circ$ latitude from the SB-ACC) and all other stations ($> 1.0^\circ$ latitude south of the SB-ACC), and the mean productivity of these two groups were compared. Krill acoustics data were high-pass filtered to produce a set of locations for back-scatter data dominated by krill targets. The mean sea surface temperature from the hull-mounted sensor was then determined for each of these locations. The null hypothesis that the back-scatter data are uniformly distributed over the range of temperature observations was tested. Underway fluorescence and sea surface temperature data were similarly compared. Fluorometric measurements of surface chlorophyll were corrected for the photoinhibition observed at irradiances greater than 150 $\mu\text{E m}^{-2} \text{ s}^{-1}$ using a calibration function derived from the ratio of HPLC-derived

chlorophyll measurements ($\mu\text{g l}^{-1}$) from surface CTD samples to the output of the fluorometer (mV) on the ship's seawater intake. This ratio increased from 0.0091 to 0.044 for irradiances between 150 and 1,200 $\mu\text{E m}^{-2} \text{ s}^{-1}$ ($r^2 = 0.52$). Underway fluorescence was corrected for diel variability associated with the daily cycle in incident irradiance, and then correlated with sea surface temperature. These underway data were used to quantify the spatial distribution of chlorophyll because of their greater resolution compared with the CTD chlorophyll data. The null hypothesis that fluorescence data are uniformly distributed over the range of temperature observations was tested.

Received 28 February; accepted 10 May 2000.

- Orsi, A. H., Whitworth III, T. & Nowlin, W. D. J. On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep Sea-Res. I* **42**, 641–673 (1995).
- Tynan, C. T. Ecological importance of the Southern Boundary of the Antarctic Circumpolar Current. *Nature* **392**, 708–7106 (1998).
- Loeb, V. *et al.* Effects of sea-ice extent and salp or krill dominance on the Antarctic food web. *Nature* **387**, 897–900 (1997).
- Nicol, S. "BROKE": a multidisciplinary survey of the waters off east Antarctica (80–150°E). *Deep-Sea Res. II* **47** (special issue), (in the press).
- Tchernia, P. & Jeannin, P. F. Circulation in Antarctic waters as revealed by iceberg tracks, 1972–1983. *Polar Record* **22**, 263–269 (1984).
- Heil, P. & Allison, I. The pattern and variability of Antarctic sea-ice drift in the Indian Ocean and Western Pacific sectors. *J. Geophys. Res.* **104**, 15789–15802 (1999).
- Mackintosh, N. A. Distribution of post-larval krill in the Antarctic. *Discovery Rep.* **36**, 1–94 (1973).
- Mizroch, S. A., Rice, D. W., Bengtson, J. L. & Larson, S. W. Preliminary atlas of Balanopterid whale distribution in the Southern Ocean based on pelagic catch data. *CCAMLR Selected Scientific Papers* **1985**, 113–193 (Commission for the Conservation of Antarctic Marine Living Resources, Hobart, Australia, 1985).
- Comiso, J. C., McClain, C. R., Sullivan, C. W., Ryan, J. P. & Leonard, C. L. Coastal zone color scanner pigment concentrations in the Southern Ocean and relationships to geophysical surface features. *J. Geophys. Res.* **98**, 2419–2451 (1993).
- Weber, L. H. & El Sayed, S. Z. in *Antarctic Nutrient Cycles and Food Webs* (eds Siegfried, W. R., Condy, P. R. & Laws, R. M.) 284–293 (Springer, Berlin, Heidelberg, 1985).
- Veit, R. R., Silverman, E. D. & Everson, I. Aggregation patterns of pelagic predators and their principal prey, Antarctic krill, near south Georgia. *J. Anim. Ecol.* **62**, 551–564 (1993).
- Nishikawa, J. *et al.* Distribution of salps near the South Shetland Islands during austral summer, 1990–1991 with special reference to krill distribution. *Polar Biol.* **15**, 31–39 (1995).
- Kawaguchi, S., Ichii, T., Naganobu, M. & de la Mare, W. K. Do krill and salps compete? Contrary evidence from the krill fisheries. *CCAMLR Sci.* **5**, 205–216 (1998).
- Hosie, G. W. in *Southern Ocean Ecology, the BIOMASS Perspective* (ed. El-Sayed, S. Z.) 93–123 (Cambridge Univ. Press, Cambridge, 1994).
- Le Fevre, J., Legendre, L. & Rivkin, R. B. Fluxes of biogenic carbon in the Southern ocean: roles of large microphagous zooplankton. *J. Mar. Systems* **17**, 325–345 (1998).
- Perissinotto, R. & Pakhomov, E. A. The trophic role of the tunicate *Salpa thompsoni* in the Antarctic marine ecosystem. *J. Mar. Systems* **17**, 361–374 (1998).
- de la Mare, W. K. Abrupt mid-twentieth century decline in Antarctic sea ice extent from whaling records. *Nature* **389**, 57–60 (1997).
- White, W. B. & Peterson, R. G. An Antarctic circumpolar wave in surface pressure, wind, temperature and sea ice extent. *Nature* **380**, 699–702 (1996).
- Priddle, J. *et al.* in *Antarctic Ocean and Resources Variability* (ed. Sahrhage, D.) 169–182 (Springer, Berlin, Heidelberg, 1988).
- Nicol, S. & Endo, Y. Krill fisheries: development, management and ecosystem implications. *Aquat. Living Resour.* **12**, 1–17 (1999).
- Jolly, G. M. & Hampton, I. A stratified random transect design for acoustic surveys of fish stocks. *Can. J. Fish. Aquat. Sci.* **47**, 1282–1291 (1990).
- Greene, C. H., Stanton, T. K., Wiebe, P. H. & McClatchie, S. Acoustic estimates of Antarctic krill. *Nature* **349**, 110 (1991).
- Thiele, D., Chester, E. T. & Gill, P. C. Cetacean distribution off eastern Antarctica (80–150°E) during the Austral summer of 1995/96. *Deep-Sea Res. II* **47**, (in the press).
- Anon. BIOMASS Working Party on Bird Biology: Recording distribution and abundance of seabirds at sea in the Southern Ocean: methods used in the BIOMASS Programme. *Mar. Ornithol.* **20**, 51–59 (1984).
- Jeffrey, S. W., Mantoura, R. F. C. & Wright, S. W. (eds) *Phytoplankton Pigments in Oceanography: Guidelines to Modern Methods* (UNESCO, Paris, 1997).
- Lewis, M. R. & Smith, J. C. A small volume, short-incubation-time method for measurement of photosynthesis as a function incident irradiance. *Mar. Ecol. Prog. Ser.* **13**, 99–102 (1983).
- Platt, T., Gallegos, C. L. & Harrison, W. G. Photoinhibition of photosynthesis in natural assemblages of marine phytoplankton. *J. Mar. Res.* **38**, 687–701 (1980).
- Mackey, D. J., Parslow, J. S., Higgins, H. W., Griffiths, F. B. & O'Sullivan, J. E. Plankton productivity and biomass in the western equatorial Pacific: biological and physical controls. *Deep-Sea Res. II* **42**, 499–533 (1995).
- Bishop, J. K. & Rossow, W. B. Spatial and temporal variability of global surface solar irradiance. *J. Geophys. Res.* **99**, 16839–16858 (1991).
- Jacka, T. J. Antarctic CRC and Australian Antarctic Division Climate Data Sets (<http://www.antarcrc.utas.edu.au/~jacka/climate.html>) (1997).
- Quetin, L. B., Ross, R. M., Frazer, T. K. & Haberman, K. L. in *Foundations for Ecological Research West of the Antarctic Peninsula* (eds Ross, R. M., Hofmann, E. E. & Quetin, L. B.) 357–371 (Antarctic Research Series 70, American Geophysical Union, Washington DC, 1996).

Acknowledgements

We would like to acknowledge the critical input of all those who took part in the BROKE voyage and to thank L. Barmuta and A. Constable for statistical advice.

Correspondence and requests for materials/information should be addressed to S.N. (e-mail: stephe_nic@antdiv.gov.au).

Diversity peaks at intermediate productivity in a laboratory microcosm

Rees Kassen*, Angus Buckling‡, Graham Bell*† & Paul B. Rainey‡

* Department of Biology, McGill University, 1205 Dr Penfield Avenue, Montreal, Quebec, Canada H3A 1B1

† Redpath Museum, McGill University, 859 Sherbrooke Street West, Montreal, Quebec, Canada H3A 2K6

‡ Department of Plant Sciences, University of Oxford, South Parks Road, Oxford OX1 3RB, UK

The species diversity of natural communities is often strongly related to their productivity. The pattern of this relationship seems to vary: diversity is known to increase monotonically with productivity, to decrease monotonically with productivity, and to be unimodally related to productivity, with maximum diversity occurring at intermediate levels of productivity^{1–3}. The mechanism underlying these patterns remains obscure, although many possibilities have been suggested^{3–6}. Here we outline a simple mechanism—involving selection in a heterogeneous environment—to explain these patterns, and test it using laboratory cultures of the bacterium *Pseudomonas fluorescens*. We grew diverse cultures over a wide range of nutrient concentrations, and found a strongly unimodal relationship between diversity and productivity in heterogeneous, but not in homogeneous, environments. Our result provides experimental evidence that the unimodal relationship often observed in natural communities can be caused by selection for specialized types in a heterogeneous environment.

Productivity is the rate of production of organic matter by a community⁷, and is in general determined by the rate of energy supply. At continental scales, the diversity of plants and animals usually increases monotonically with productivity, or some suitable surrogate measure^{3,8}. At the smaller scale of regions within continents, the patterns are more complex. The prevailing view has been that diversity and productivity are unimodally related^{9–13}, such that diversity at first increases with productivity or production, reaches a maximum for intermediate values, and then declines to lower levels in highly productive systems. Support for this view comes from nutrient addition experiments in terrestrial plant communities and aquatic systems^{14–19}, although the generality of the unimodal pattern has been questioned and the theory criticized^{3,6}.

The mechanism underlying this pattern remains unclear, although it is likely to involve competition in heterogeneous environments. Diversity can be maintained if relative competitive ability, or fitness, varies among niches within an environment²⁰, where 'niches' is usually interpreted as a collection of sites offering different conditions of growth. This will create antagonistic selection that preserves local adaptation, provided that each niche is capable of supporting similar numbers of individuals in the community²¹. If niches are unable to support approximately equal numbers of individuals, then the type that is best-adapted to the most productive niche will come to dominate the community^{22,23} and diversity will be low.

The effect of varying productivity can be described by a simple modification of Levene's²⁰ model for the maintenance of diversity in a heterogeneous environment. In the classic version of the model two (or more) types, which may be species or asexual genotypes, are distributed at random into two (or more) niches. Selection occurs within each niche, followed by reproduction. Each niche then contributes a fixed proportion of offspring c_j , such that $\sum c_j = 1$, to

a common pool, and the sequence is repeated. With just two types, the frequency of type 1 at equilibrium is

$$p_1^* = \frac{[c_1(w_{11}w_{22} - w_{12}w_{21}) + w_{21}(w_{12} - w_{22})]}{(w_{21} - w_{11})(w_{12} - w_{22})} \quad (1)$$

where w_{ij} is the fitness of the i th type in the j th niche, and both types coexist if $0 < p_1^* < 1$ (for a full derivation, see Box 1). Note that two conditions must be met to ensure coexistence of both types. The first is the relative fitness of the types within each niche: each type must be superior in one niche but not the other, giving rise to antagonistic selection. This will be the case when the term describing genotype-by-environment interaction for fitness, $(w_{11}w_{22} - w_{12}w_{21})$, is non-zero. The second is the fraction of the total population contributed by each niche, c_j . Each niche must contribute approximately equal proportions to the total population. When both conditions hold, the equilibrium point is stable and diversity is maintained through negative frequency-dependent selection; either type has greater fitness when rare because it is able to invade the niche to which it is better adapted.

Productivity can be incorporated into the classic Levene model by allowing absolute fitness, w_{ij} , to be a function of resource supply rate, R . The Michaelis–Menton relation provides a suitable functional form:

$$w_{ij} = \frac{w_{\max,ij}(R - R_{\min,ij})}{K_{m,ij} + (R - R_{\min,ij})} \quad (2)$$

where, for each type i in niche j , $R_{\min}$ is the minimal resource supply rate required to maintain viability, $w_{\max}$ is the maximum number of individuals that could be produced at maximal resource supply rates, and K_m , the resource supply rate at $w_{\max}/2$, is a measure of the ability to convert resources into individuals at half-maximal resource supply rates. These parameters thus determine the number of individuals produced by a given type within a niche across the entire range of resource supply rates. We assume, as in the original model, that relative fitness within a niche remains constant across the entire range of resource supply rates. Given antagonistic selection, the outcome of selection is then governed solely by the contribution of each niche to the total population, which responds to resource supply rate as specified in equation (2).

To see this more clearly, consider Figs 1a and 2a, which depict two sorts of relationship between fitness and resource supply rate for two types, each specialized to a different niche (and therefore satisfying the necessary condition that selection be antagonistic). In both cases, the community is dominated at low resource supply rates by the type that is fittest in the niche where growth requires the lowest minimum resource supply rate (Figs 1a and 2a). As resource supply rate increases, the production of the two niches becomes more nearly equal and diversity can be maintained. At high resource supply rates, however, the production of the niches diverge, one contributing a larger fraction of the total population than the other. The type that is best adapted to the more productive niche then prevails, and diversity is lost. This generates a unimodal relationship between diversity and resource supply rates (Figs 1b and 2b). Note that the type dominating the community at high productivity depends on how fitness varies with resource supply rate (Figs 1c and 2c).

We note that the results of our model do not seem to be specific to the set of parameter values that we have chosen or to the use of the Michaelis–Menton function. Any set of curves that cross once within the range of resource supply rates and that share the same qualitative form as the Michaelis–Menton function will probably lead to similar results. We have found that it is possible to obtain more complex patterns, such as a bimodal relationship between diversity and productivity, if the curves for each niche cross more than once (data not shown). Nevertheless, such complex patterns have not been observed in nature³ and so we do not consider them further.

Our theory thus represents a simple and general explanation for unimodal diversity–productivity curves at a single trophic level. Note that it does not require a shift from resource limitation to light limitation (or any other single limiting factor) along a productivity gradient^{1,5,6} or predation^{24,25} to generate the unimodal relationship. Its key features are simply niche specialization, leading to changes in the ranking of relative fitness across niches, and the relative rates of production of different niches, both of which are readily amenable to experimental manipulation using microbial systems. Microbes are well-suited for studying community ecology and evolution; their small size and rapid generation time allow large populations to be maintained under defined conditions. Bacterial populations have the further advantage of being asexual, meaning that each genotypic lineage corresponds to a species in a sexual community, and we can thus test theories of species diversity by the response of genotypic diversity to experimental manipulations.

We used cultures of *Pseudomonas fluorescens* which undergo rapid adaptive radiation producing a variety of genotypes specialized to different ecological niches (for example, the surface, liquid phase or walls of the culture vessel) that are easily identified by their colony morphology when grown on agar plates²⁶. Diversity is maintained through negative frequency-dependent selection in

spatially structured microcosms²⁶. When the spatial structure is removed by shaking, diversity is rapidly lost from previously diverse cultures and fails to become established in genetically uniform cultures. To create a gradient in productivity we used serial dilutions of a standard medium to which the base population is well-adapted, creating environments differing by up to three orders of magnitude in nutrient concentration. In this way, we can contrast the effects of productivity on diversity in homogeneous and heterogeneous environments.

Our model predicts a unimodal relationship between diversity and productivity in heterogeneous environments. In homogeneous environments, the model predicts little or no relationship between diversity and nutrient concentration. Further, the model predicts that the community should be dominated by a single type at both very low and very high nutrient concentrations, implying that diversity at these extremes should be similar in heterogeneous and homogeneous environments.

We obtained two highly diverse populations, each constituting the base of a replicate experiment, by allowing cultures to proliferate in unshaken microcosms containing standard KB medium for 4 days. These were used to inoculate tubes containing medium that varied in concentration over the range of factors between 1/128 and 8 times that of standard. After two days of growth the experimental cultures were vortexed and plated to enable colony

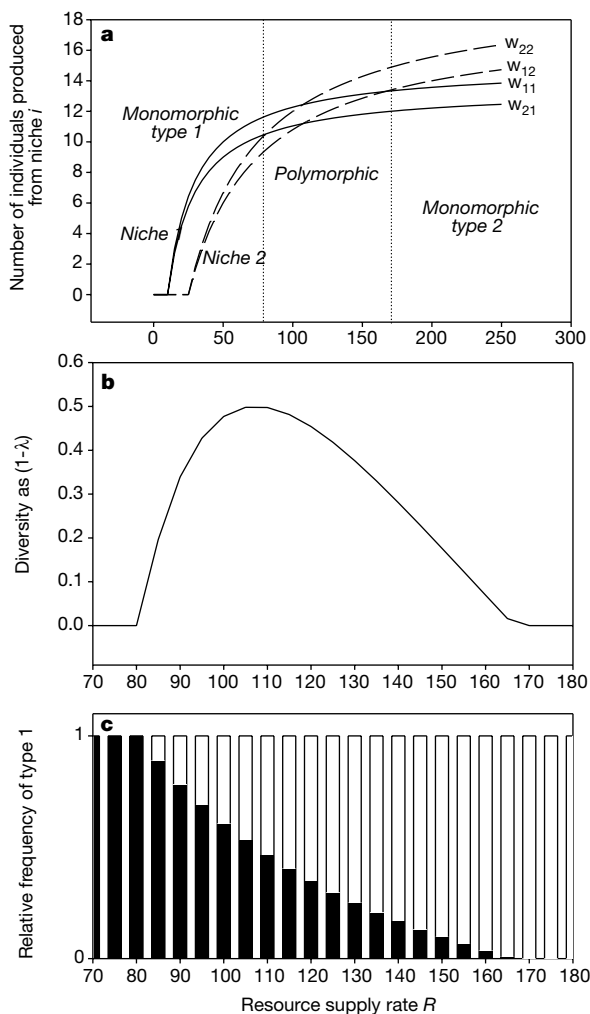


Figure 1 Maintenance of two types in two ecological niches over a range of productivities. **a**, Theoretical curves, explained in text. Parameter values are type 1 in niche 1, $R_{\min} = 10$, $K_m = 20$, $w_{\max} = 15$; type 2 in niche 2, $R_{\min} = 25$, $K_m = 50$, $w_{\max} = 20$. The less-fit type in each niche has a fitness equal to 90% of that of the more-fit type at all resource supply rates. **b**, The behaviour of the diversity index $1 - \lambda$, where $\lambda = \sum p_i^2$ is Simpson's index. **c**, Relative frequencies of the two types.

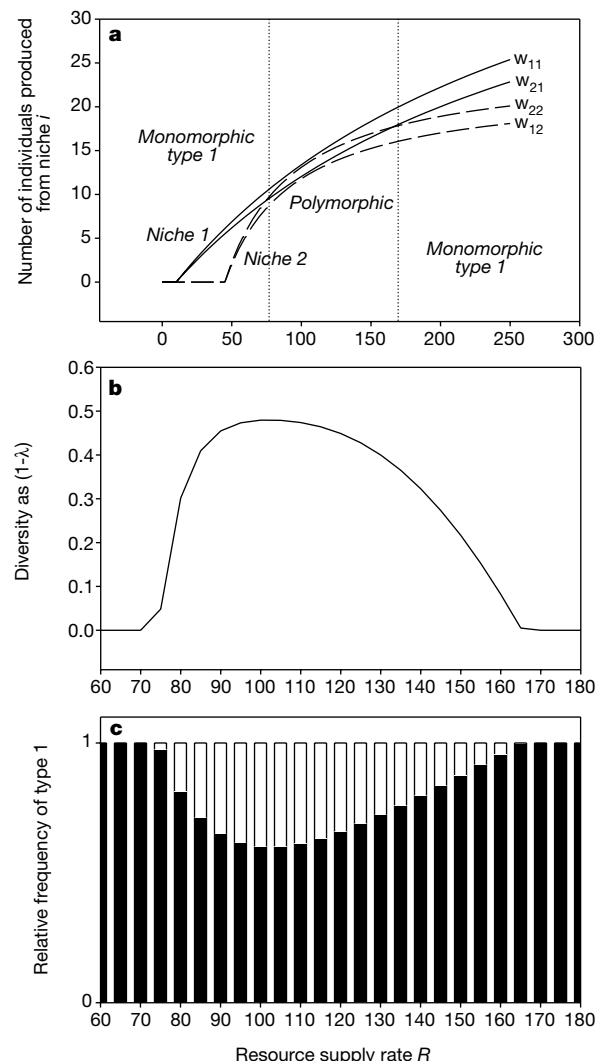


Figure 2 As in Fig. 1 but with different parameter values. Type 1 in niche 1, $R_{\min} = 10$, $K_m = 280$, $w_{\max} = 55$; type 2 in niche 2, $R_{\min} = 45$, $K_m = 50$, $w_{\max} = 25$.

morphotypes to be scored. The three basic types appearing in *P. fluorescens* cultures are 'smooth', the ancestral type which lives chiefly in the broth phase; 'fuzzy-spreader', which occupies the bottom of the vessel; and 'wrinkly-spreader', which colonizes the air-broth interface. There is considerable variation within each of these categories, and we distinguished about 10 types during this experiment.

The results, shown in Figs 3 and 4, are striking: diversity in the heterogeneous (unshaken) environments is clearly unimodally related to productivity, while in the homogeneous (shaken) environments it is not. In shaken cultures there was only a weak response of diversity to increased nutrient concentration (Figs 3 and 4, Table 1). This was significant in one case ($P = 0.032$) although not in the other ($P = 0.12$); combining the two yields $\chi^2 = 11.1$, d.f. = 4, $P \approx 0.025$. The same procedure yields $\chi^2 = 5.79$, d.f. = 4, $P = 0.3$ for the quadratic effect. In short, there was a weak and inconsistent tendency for diversity to increase monotonically with productivity in the homogeneous environments. It is conceivable that such an increase could arise if increasing nutrient concentration is associated with increasing population size, and so a larger sample of the total range variation. However the large population sizes in our experiment (at least $\sim 10^6$ – 10^7 per culture) make this an unlikely explanation. In the unshaken cultures there was again some tendency for an overall increase of diversity with productivity; this was

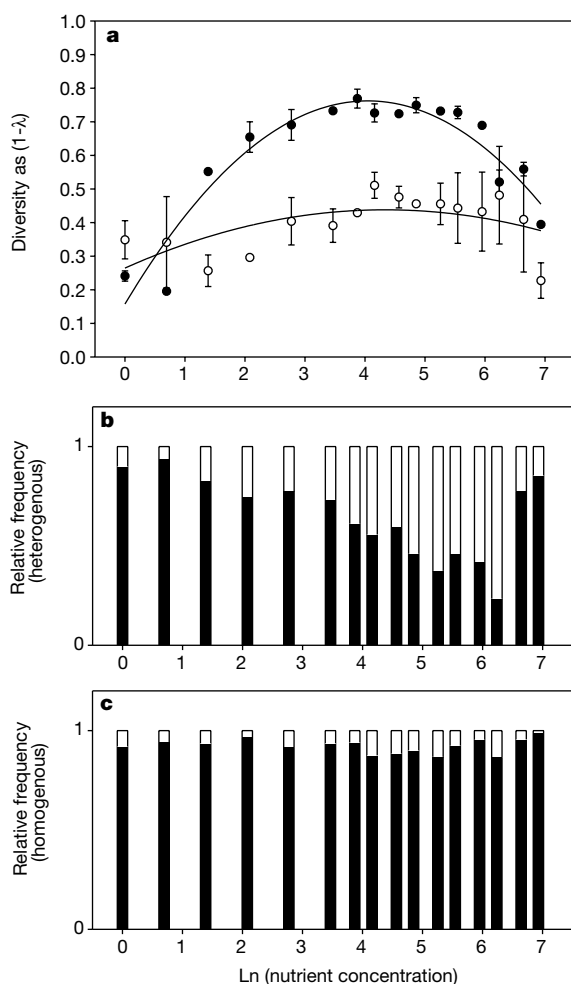


Figure 3 Response of diversity to nutrient concentration in base population 1. **a**, Diversity expressed as $1 - \lambda$. Open circles are homogeneous, and solid circles heterogeneous, cultures, with bars marking ± 1 s.e. of two replicates. **b**, Relative frequency of different colony morphotypes in the heterogeneous environment; 'smooths' (solid bars) versus 'wrinkly-spreaders' and 'fuzzy-spreaders' (open bars). **c**, As in **b**, for the homogeneous environment.

Table 1 Analysis of covariance of the experiments

Structure	Effect	d.f.	MS	F	P	r^2
First base population						
Unshaken	Linear	1	0.088	21.6	<0.0005	0.18
	Quadratic	1	0.357	87.7	<0.0001	0.89
	Error	13	0.004			
Shaken	Linear	1	0.014	2.77	0.12	0.14
	Quadratic	1	0.022	4.41	0.06	0.35
	Error	13	0.005			
Second base population						
Unshaken	Linear	1	0.014	4.34	0.06	0.08
	Quadratic	1	0.122	38.8	<0.0001	0.76
	Error	13	0.003			
Shaken	Linear	1	0.065	5.77	0.03	0.31
	Quadratic	1	0.00	0.00	0.99	0.31
	Error	13	0.011			

The response variable is $1 - \lambda$ as shown in Fig. 1.

somewhat stronger than in homogeneous environments (combined $\chi^2 = 20.9$, d.f. = 4, $P < 0.001$), but diversity in homogeneous and heterogeneous environments was very similar, both at very low and at very high productivity.

The most remarkable feature of the results in the heterogeneous environments, however, was the negative quadratic effect, which was highly significant ($P < 0.0001$) in both replicates (Table 1). This reflects the modal relationship between diversity and productivity, with diversity in the heterogeneous environment exceeding that in the homogeneous environment and peaking at almost the same intermediate concentration in both cases (Figs 3 and 4). Moreover, our model predicts that competition will lead to a sequential

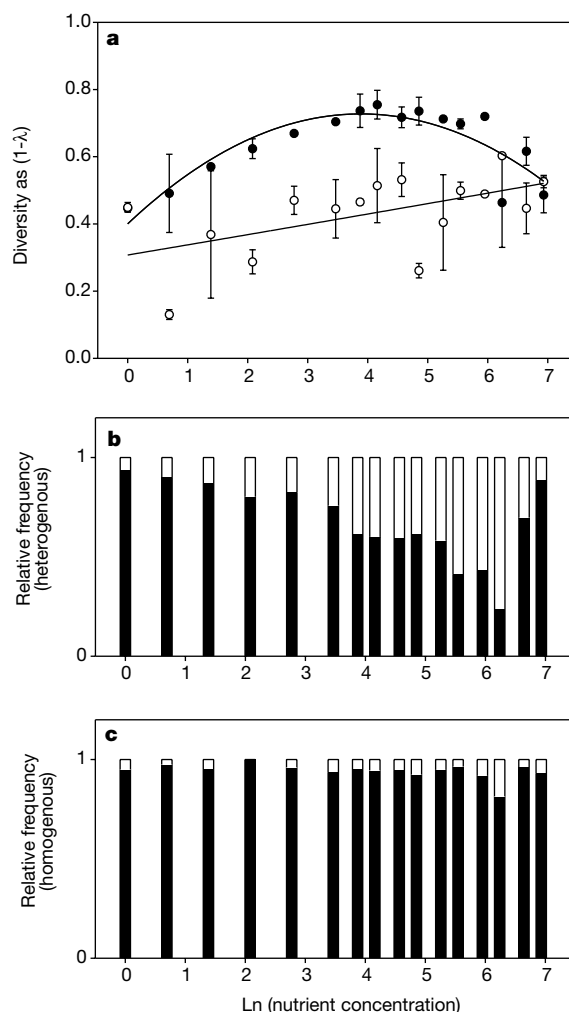


Figure 4 Response of diversity to nutrient concentration in base population 2, as in Fig. 3.

replacement at intermediate productivity of one type by the other (Figs 1c and 2c), which is what we observed in our experiment (Figs 3b and 4b).

These results are broadly consistent with the simple model described in Figs 1 and 2, and mirror the relationships between diversity and productivity or production that have been repeatedly observed at small spatial scales in natural communities. Moreover, the mechanism by which diversity is maintained in this system through negative frequency-dependent selection is well-understood: at standard nutrient concentrations, a rare type will invade when its niche is available. In our experiment, the two most common categories were 'smooth' and 'wrinkly-spreader'. The former occupies the broth phase and was dominant at low nutrient concentration; the latter occupies the surface, and its success depends on the formation of a coherent mat. At low nutrient concentration there are too few cells to make this possible, but at higher concentrations, and thus higher overall population density, 'wrinkly-spreader' can invade and diversity increases. At the highest concentrations, the mat formed by 'wrinkly-spreader' is still visible, but 'smooth' appears to be the superior competitor, a result consistent

with the predictions of the model outlined in Fig. 2. In natural communities, of course, the range of ecological opportunities will be much greater and the degree of specialization much finer than in our simplified laboratory system. Nevertheless, our experiment supports the interpretation of the modal diversity-productivity relationship as arising from selection for niche specialization in a heterogeneous environment.

At larger spatial scales it has been observed that diversity tends to increase linearly with productivity^{3,8}, and our model suggests two reasons for this. First, different kinds of niches will be included as the extent of the environment increases, leading to an overall increase in the environmental variance of fitness²⁷ and likewise an increase in the quantity of genotype-environment interaction for fitness²⁸. Second, immigration will also decrease as the distance between niches becomes larger relative to dispersal distance. The combined effects of stronger selection and lower migration means that local adaptation, and so diversity, can be maintained more readily. Over larger spatial scales, then, the linear relationship between diversity and productivity should be a function of the grain of the environment, relative to the size and dispersal ability of the organisms concerned. □

Box 1

The simplest version of the Levene model

Two asexual types are distributed at random to two niches. All individuals initially present are equally fecund, but the genotypes may survive differently in the two niches. (The same conclusions follow if fecundity rather than survival be the target of selection.) The sequence of events is settlement followed by differential survival, then reproduction, after which the parents die and a fixed number of offspring from each niche form a dispersal pool from which they are distributed at random between the niches. Let w_{ij} signify the fitness (probability of survival) of the i th genotype in the j th niche, which contributes a fraction c_j of the dispersal pool. If the initial frequency of type 1 is p , then its frequency p' after a single round of selection will be

$$p' = \left[\frac{pw_{11}}{pw_{11} + (1-p)w_{21}} \right] c_1 + \left[\frac{pw_{12}}{pw_{12} + (1-p)w_{22}} \right] (1 - c_1)$$

Setting $\Delta p = p' - p = 0$, we can solve for the equilibrium frequency of type 1, p^* :

$$p^* = \frac{[c_1(w_{11}w_{22} - w_{12}w_{21}) + w_{21}(w_{12} - w_{22})]}{(w_{21} - w_{11})(w_{12} - w_{22})}$$

which is equation (1) in the text. This expression has three parts. The denominator is a genetic variance, and the right-hand term in the numerator specifies the genotype referred to. The left-hand term in the numerator contains the two crucial elements of the model. The first is the fixed contribution of a niche to the dispersal pool, c_i . This constitutes local density regulation, which creates negative frequency-dependent selection because the probability that a surviving individual of a given type will be included in the dispersal pool will be inversely proportional to the initial frequency of that type in its niche. The second element is genotype-environment interaction. If there is no multiplicative genotype-environment interaction, relative fitness is the same in both niches, so that $w_{11}/w_{21} = w_{12}/w_{22}$, and consequently $w_{11}w_{22} - w_{12}w_{21} = 0$. The term is non-zero only if relative fitness differs between niches. A necessary condition for $0 < p^* < 1$ is that the ranking of relative fitness changes between niches, that is, $w_{11} > w_{21}$ and $w_{22} > w_{12}$ (or the reverse), so that each type is superior in one of the niches and inferior in the other. Sufficient conditions for the maintenance of diversity involve both the intensity of niche-specific selection and the balance of niche-specific production. Setting $w_{12} = w_{21} = 1$ for clarity, these conditions are

$$w_{11} > \frac{[c_1 - (1 - w_{22})]}{c_1 w_{22}} \quad \text{and} \quad w_{22} > \frac{(w_{11} - c_1)}{w_{11}(1 - c_1)}$$

implying that as the share of the dispersal pool contributed by a given niche decreases, the greater the difference in relative fitness in that niche required to maintain diversity.

Methods

Base populations

Each base population was obtained by culturing *Pseudomonas fluorescens* isolate SBW25 in 6 ml of King's B (KB) medium for 4 days.

Experimental procedure

Samples from the two base populations were washed three times in 10 mM phosphate buffer and then starved for two hours. Experimental cultures were then inoculated with about 10^3 cells. Experimental culture media were mixtures of KB nutrients (glycerol and proteose peptone) and M9 mineral salts, made up by serial dilution in M9 salts to concentrations of $8 \times$, ..., $1/128 \times$ standard KB nutrient concentration. After 48 h, cultures had reached 10^8 – 10^9 cells and were stored in 50% glycerol at -80°C . Diversity was estimated by plating the 48-h cultures onto KB agar and scoring the morphology of approximately 100 random colonies.

Statistical procedure

Diversity was expressed as the complement of Simpson's index²⁹, $1 - \lambda = 1 - \sum p_i^2 = \sum p_i(1 - p_i)$ where p_i is the frequency of the i th type. Expressing diversity as richness, or the number of types, gives the same pattern (data not shown).

Received 25 April; accepted 1 June 2000.

1. Tilman, D. *Resource Competition and Community Structure* (Princeton Univ. Press, Princeton, New Jersey, 1982).
2. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, UK, 1995).
3. Waide, R. B. *et al.* The relationship between productivity and species richness. *Annu. Rev. Ecol. Syst.* **30**, 257–300 (1999).
4. Rosenzweig, M. L. & Abramsky, Z. in *Species Diversity in Ecological Communities* (eds Ricklefs, R. & Schluter, D.) 52–65 (Univ. Chicago Press, Chicago, Illinois, 1993).
5. Tilman, D. & Pacala, S. in *Species Diversity in Ecological Communities* (eds Ricklefs, R. & Schluter, D.) 13–25 (Univ. Chicago Press, Chicago, Illinois, 1993).
6. Abrams, P. A. Monotonic or unimodal diversity-productivity gradients: what does competition theory predict? *Ecology* **76**, 2019–2027 (1995).
7. Odum, E. P. *Fundamentals of Ecology* (Saunders, Philadelphia, 1959).
8. Currie, D. J. Energy and large-scale patterns of animal- and plant-species richness. *Am. Nat.* **137**, 27–49 (1991).
9. Abramsky, Z. & Rosenzweig, M. L. Tilman's predicted productivity-diversity relationship shown by desert rodents. *Nature* **309**, 150–151 (1984).
10. Wisheu, I. C. & Keddy, P. A. Species richness-standing crop relationships along four lakeshore gradients: constraints on the general model. *Can. J. Bot.* **67**, 1609–1617 (1989).
11. Shipley, B., Keddy, P. A., Gaudet, C. & Moore, D. R. J. A model of species density in shoreline vegetation. *Ecology* **72**, 1658–1667 (1991).
12. Moore, D. R. J. & Keddy, P. A. The relationship between species richness and standing crop in wetlands: the importance of scale. *Vegetatio* **79**, 99–106 (1989).
13. Wheeler, B. D. & Shaw, S. C. Above-ground crop mass and species richness of the principal types of herbaceous rich-fen vegetation of lowland England and Wales. *J. Ecol.* **79**, 285–301 (1991).
14. Bakelaar, R. G. & Odum, E. P. Community and population level responses to fertilization in an old-field ecosystem. *Ecology* **59**, 660–665 (1978).
15. Goldberg, D. E. & Miller, T. E. Effects of different resource additions on species diversity in an annual plant community. *Ecology* **71**, 213–225 (1990).
16. Wilson, S. D. & Tilman, D. Components of plant competition along an experimental gradient of nitrogen availability. *Ecology* **72**, 1050–1065 (1991).
17. Al-Mufti, M. M. *et al.* A quantitative analysis of shoot phenology and dominance in herbaceous vegetation. *J. Ecol.* **65**, 759–791 (1977).

18. Yount, J. L. Factors that control species numbers in Silver Springs, Florida. *Limnol. Oceanogr.* **1**, 286–295 (1956).
19. Schindler, D. W. Experimental perturbation of whole lakes as tests of hypotheses concerning ecosystem structure and function. *Oikos* **57**, 25–41 (1990).
20. Levene, H. Genetic equilibrium when more than one ecological niche is available. *Am. Nat.* **87**, 331–333 (1953).
21. Maynard Smith, J. & Hoekstra, R. Polymorphism in a varied environment: how robust are the models? *Genet. Res.* **35**, 45–57 (1980).
22. Dempster, E. R. Maintenance of genetic heterogeneity. *Cold Spring Harb. Symp. Quant. Biol.* **20**, 213–224 (1955).
23. Grime, J. P. Control of species density in herbaceous vegetation. *J. Environ. Mgmt* **1**, 151–167 (1973).
24. Leibold, M. A. A graphical model of keystone predators in food webs: trophic regulation of abundance, incidence, and diversity patterns in communities. *Am. Nat.* **147**, 784–812 (1996).
25. Holt, R., Grover, J. & Tilman, D. Simple rules for interspecific dominance in systems with exploitative and apparent competition. *Am. Nat.* **144**, 741–771 (1994).
26. Rainey, P. B. & Travisano, M. Adaptive radiation in a heterogeneous environment. *Nature* **394**, 69–72 (1998).
27. Bell, G. The ecology and genetics of fitness in *Chlamydomonas*. V. The relationship between genetic correlation and environmental variance. *Evolution* **46**, 561–566 (1992).
28. Bell, G. *et al.* The spatial structure of the physical environment. *Oecologia* **96**, 114–121 (1993).
29. Simpson, E. H. Measurement of diversity. *Nature* **163**, 688 (1949).

Acknowledgements

We thank K. McCann and D. Schoen for comments on an earlier version of the manuscript. This work was supported by the Canadian NSERC (G.B. and R.K.), the UK BBSRC and NERC (P.B.R. and A.B.) and the British Council.

Correspondence and requests for materials should be addressed to R.K. (e-mail: rkassen@bio1.lan.mcgill.ca).

Herbivory-induced volatiles elicit defence genes in lima bean leaves

Gen-ichiro Arimura^{*†}, Rika Ozawa^{*†}, Takeshi Shimoda^{*†}, Takaaki Nishioka[‡], Wilhelm Boland[§] & Junji Takabayashi[†]

^{*} Bio-oriented Technology Research Advancement Institution, Tokyo 105-0001, Japan

[†] Laboratory of Ecological Information, Graduate School of Agriculture, Kyoto University, Kitashirakawa, Kyoto 606-8502, Japan

[‡] Laboratory of Insect Physiology, Graduate School of Agriculture, Kyoto University, Kitashirakawa, Kyoto 606-8502, Japan

[§] Max Planck Institute for Chemical Ecology, Carl-Zeiss-Promenade 10, D-07745 Jena, Germany

In response to herbivore damage, several plant species emit volatiles that attract natural predators of the attacking herbivores^{1–5}. Using spider mites (*Tetranychus urticae*) and predatory mites (*Phytoseiulus persimilis*)^{1–4}, it has been shown that not only the attacked plant but also neighbouring plants are affected, becoming more attractive to predatory mites^{3,6} and less susceptible to spider mites⁶. The mechanism involved in such interactions, however, remains elusive. Here we show that uninfested lima bean leaves activate five separate defence genes when exposed to volatiles from conspecific leaves infested with *T. urticae*, but not when exposed to volatiles from artificially wounded leaves. The expression pattern of these genes is similar to that produced by exposure to jasmonic acid. At least three terpenoids in the volatiles are responsible for this gene activation; they are released in response to herbivory but not artificial wounding. Expression of these genes requires calcium influx and protein phosphorylation/dephosphorylation.

We placed two lima bean (*Phaseolus lunatus* cv. Sieva) leaves in a plastic cup. Each of the leaves was infested with 100 female *T. urticae*. Within a lidded glass container, four to six lima bean leaves in a vial were kept together with the *T. urticae*-infested leaves in a plastic cup

for 1 or 3 days. The *T. urticae*-infested leaves emitted the '*T. urticae*-induced' volatiles, and the leaves in the vial received the volatiles. We analysed the expression patterns of six defence genes in the leaves emitting the *T. urticae*-induced volatiles and in the leaves receiving the volatiles, and identified the volatile substances that are responsible for eliciting defence genes in the receiver leaves.

In the *T. urticae*-infested leaves, we detected the transcripts of six defence genes: genes for the basic type of pathogen-related (PR) proteins (PR-2 (β -1,3-glucanase) and PR-3 (chitinase)); an acidic type of PR-4 (chitinase); lipoxygenase (LOX) in the octadecanoid pathway; phenylalanine ammonia-lyase (PAL) in the phenylpropanoid pathway; and farnesyl pyrophosphate synthetase (FPS) related to isoprene biosynthesis (Fig. 1a). Artificial wounding of lima bean leaves elicited four of the defence genes (the acidic PR-4 and FPS genes were not elicited). We have proposed a signalling mechanism such that infestation by *T. urticae* activates both jasmonic acid (JA)- and salicylic acid (SA)-signalling pathways in lima bean leaves⁷. These compounds are known as active components of wound^{8–10} or pathogen^{10,11} response in higher plants. Exogenous JA or methyl salicylate (MeSA) reproduced the expression patterns of six defence genes observed in the *T. urticae*-infested leaves (Fig. 1a). The present results constitute further evidence supporting the proposed signalling mechanism.

In lima bean leaves exposed to the *T. urticae*-infested leaf volatiles for one day in the same glass container, we also detected the transcripts of five out of the six defence genes (Fig. 1a). Transcripts of the basic PR-3 and PAL genes decreased after three days. In a control experiment, lima bean leaves received the 'wound-induced' volatiles that lima bean leaves emit in response to artificial wounding. We detected only the transcript of a basic PR-2 gene in the receiver leaves. Neighbouring lima bean plants can respond to the *T. urticae*-induced and wound-induced volatiles respectively by eliciting different combinations of defence genes.

LOX in the octadecanoid pathway is a key enzyme in the biosynthesis of JA¹². Enzymatic activity of LOX in lima bean leaves that received the *T. urticae*-induced volatiles increased as high as the activity in the *T. urticae*-infested leaves (Table 1). Previous treatment with a LOX inhibitor, SHAM, completely blocked the expression of basic PR-2, basic PR-3, LOX, PAL and FPS genes in the receiver leaves (Fig. 1b). This inhibitory effect was overcome by simultaneous treatment of exogenous JA with the inhibitor. In the receiver leaves, the *T. urticae*-induced volatiles might induce *de novo* synthesis of LOX followed by biosynthesis of JA, which subsequently activates a subset of defence genes.

Treatment with a chelator of extracellular Ca²⁺ ion, BAPTA, or an inhibitor of serine/threonine protein kinases, staurosporine, completely suppressed the expression of the five defence genes in lima bean leaves that received the *T. urticae*-induced volatiles (Fig. 1b). An inhibitor of type 1 and/or type 2A protein phosphatases, calyculin A, suppressed the expression of basic PR-3, PAL and FPS genes completely, but that of basic PR-2 and LOX genes only slightly. Therefore, the signalling pathway(s) mediating expression of defence genes in the receiver leaves would include the calcium influx into cells, protein phosphorylation, and dephosphorylation steps.

Table 1 Inductions of LOX activity in lima bean leaves

Duration (days)	Enzyme activity (relative activity*)		
	Control	Emitter leaves infested by <i>T. urticae</i>	Receiver leaves
0	1.43 ± 0.24 (1.00)	–	–
1	–	2.18 ± 0.33 (1.52)	6.95 ± 1.55 (4.87)
3	1.29 ± 0.19 (0.94)	5.08 ± 1.41 (3.57)	3.80 ± 0.35 (2.66)

*Relative to the activity of the control experiment at day 0. Lima bean leaves were either infested with *T. urticae* or receiving the *T. urticae*-induced volatiles. Values are means ± standard deviation of three experiments (ΔA_{234} per mg protein per min).

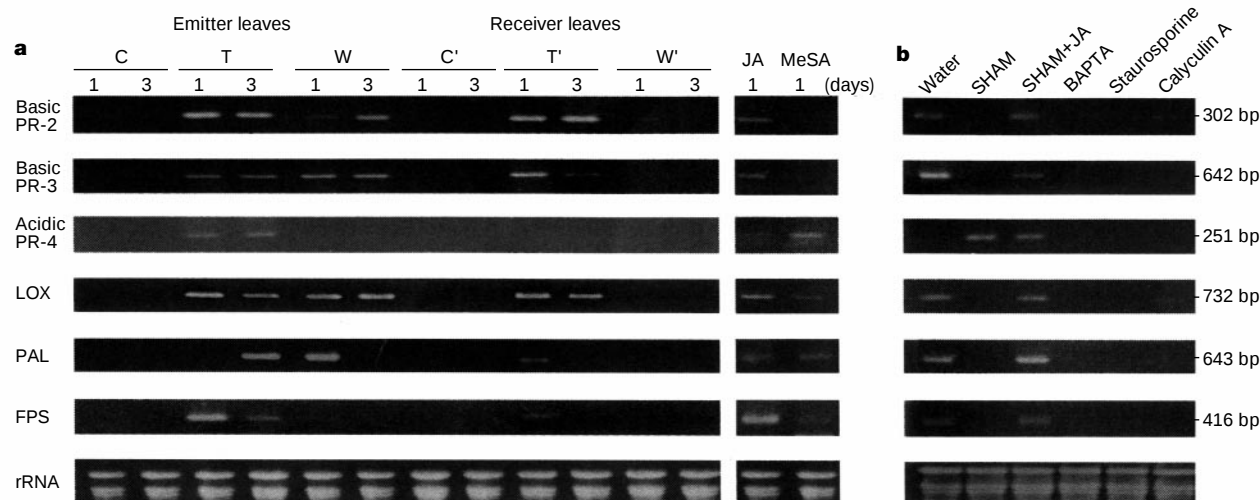


Figure 1 Expression of defence genes in lima bean leaves. **a**, Receiver leaves (C', T' and W') were exposed to volatiles from emitter leaves with the same symbols: uninfested (C), *T. urticae*-infested (T) or artificially wounded (W) leaves. An aqueous jasmonic acid (0.1 mM; Wako, Japan) was applied from the petioles of leaflets (JA). Methyl salicylate (Wako) was applied to leaves as a gas (about 30 p.p.b.) in a glass bottle (2 litre) (MeSA). **b**, Leaves were treated with salicylhydroxamic acid (SHAM, 10 mM; Sigma, USA),

SHAM + JA (0.1 mM), 1,2-bis-(2-aminophenoxy)ethane-*N,N,N',N'*-tetraacetic acid (BAPTA, 5 mM; Sigma), staurosporine (2 μ M; Wako, Japan), or calyculin A (3 μ M; Sigma), through the petioles of leaflets. Three hours after each application, the leaves were placed in a glass container together with *T. urticae*-infested leaves for one day. bp, base pairs.

Major chemical components from the *T. urticae*-infested leaf volatiles are (*E*)- β -ocimene, (*E*)-4,8-dimethyl-1,3,7-nonatriene (DMNT), and (*E,E*)-4,8,12-trimethyl-1,3,7,11-tridecatetraene (TMTT) (Fig. 2). Concentrations of these substances increased with the gradual infestation by *T. urticae* over 24 h. In the volatiles, we did not observe enhancement in the emission of linalool and MeSA which had been reported as chemical components in *T. urticae*-induced lima bean leaf volatiles⁴. The major chemical component of the wound-induced volatile was (*Z*)-3-hexenol. Emission of this substance was transient. Terpenoids and MeSA were scarcely detected in the wound-induced volatiles for 24 h after the wound was made. Such temporal, quantitative and qualitative differences in chemical components between the *T. urticae*-induced and the wound-induced volatiles might induce the different expression patterns of defence genes in receiver leaves.

Airborne methyl jasmonate (MeJA) and MeSA were proposed to mediate plant–plant interactions in some specific cases^{8,13}. In the *T. urticae*-induced volatiles, however, we detected no MeJA but 2.97 ± 0.42 μ g MeSA per gram fresh weight per hour. This was as low as 2.71 ± 0.48 μ g MeSA per gram fresh weight per hour in the volatiles from the uninfested lima bean leaves in a control experiment. In addition, no transcript for an acidic PR-4 gene associated with MeSA was detected in the receiver leaves (see Fig. 1). Thus, MeSA is detectable in the *T. urticae*-induced volatiles but is less likely to be an active component in mediating the plant–plant interactions in the present study.

We exposed uninfested lima bean leaves to the vapours of chemically synthesized preparations of four chemicals and analysed transcripts of defence genes in the leaves (Fig. 3). DMNT and TMTT induced the expression of basic PR-2 and PR-3, LOX, PAL and FPS genes in the leaves within 3 h of exposure. Likewise, β -ocimene induced the expression of basic PR-2 and PR-3, PAL and FPS genes in 24 h. Linalool, in contrast, induced no expression for the six defence genes we analysed here. DMNT, TMTT, and β -ocimene in the *T. urticae*-induced volatiles are responsible for the induction of defensive responses in the receiver leaves. Another potential ingredient that is not detectable by the present analytical method is ethylene. Tobacco plants emit ethylene in response to herbivory¹⁴. Another two potential ingredients are (*Z*)-3-hexenol and

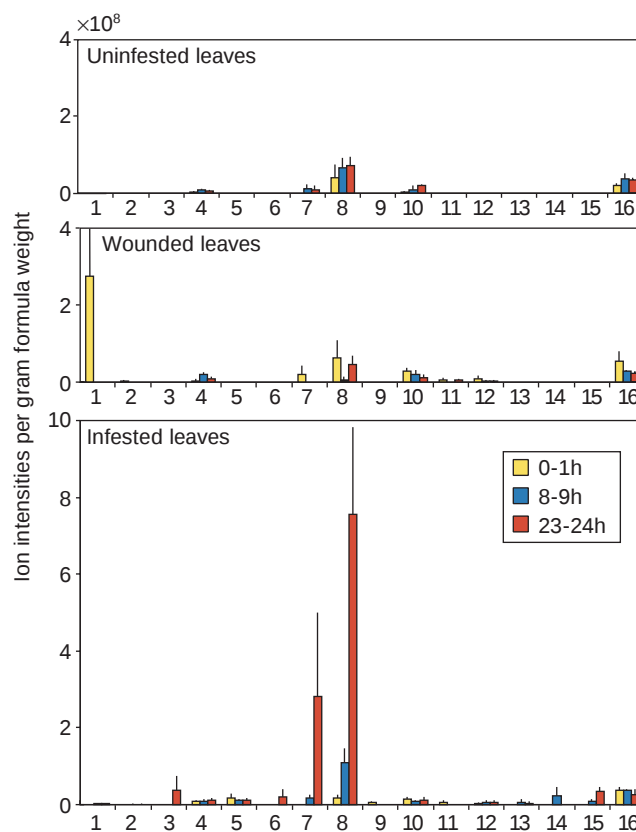


Figure 2 Compounds identified in headspace volatiles of detached leaves. The two leaves in each treatment were placed in a lidded glass container for 24 h, except for when volatile collections were performed at 0–1 h, 8–9 h and 23–24 h of the experimental period. The collected volatiles were subjected to GC-MS analysis ($n = 3$). 1, (*Z*)-3-hexenol; 2, (*E*)-2-hexenal; 3, (*Z*)-3-hexenyl acetate; 4, α -pinene; 5, limonene; 6, (*Z*)- β -ocimene; 7, (*E*)- β -ocimene; 8, DMNT; 9, menthol; 10, α -copaene; 11, junipene; 12, β -caryophyllene; 13, α -humulene; 14, germacrene d; 15, TMTT; 16, methyl salicylate.

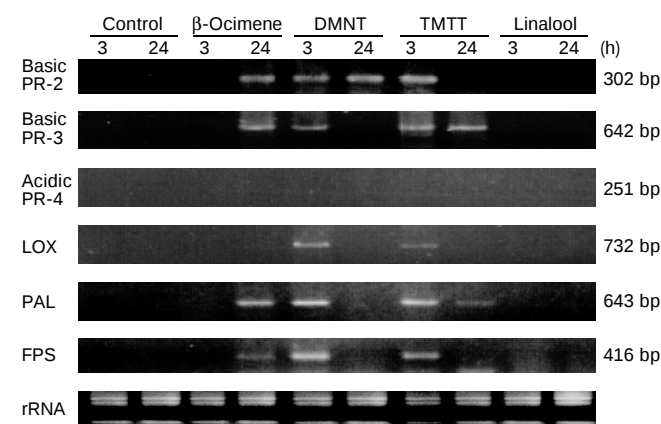


Figure 3 Expression of genes in leaves fumigated with a terpenoid compound. We enclosed unfested leaves in a glass container together with a piece of cotton wool containing a chemical (β -ocimene, DMNT, TMTT, and linalool) dissolved in dichloromethane for 3 or 24 h. A piece of cotton wool containing dichloromethane was the control.

(*Z*)-3-hexenyl acetate. These are responsible for the defensive responses in *Arabidopsis* plants¹⁵.

We evaluated how suitable the lima bean leaves receiving the *T. urticae*-induced volatiles are as a food resource for *T. urticae*. We exposed lima bean leaves to the *T. urticae*-induced volatiles for three days then infested them with 100 female *T. urticae* per leaf for another three days ('sample' leaves). We measured the loss of chlorophyll in the infested leaves as an indication of consumption by *T. urticae*¹⁶. In the control experiment, we exposed lima bean leaves to the volatiles from unfested leaves then infested them with the same number of female *T. urticae* ('control' leaves). The amounts of chlorophyll in the sample and the control leaves were 0.56 ± 0.16 and 0.56 ± 0.01 mg per gram fresh weight without infestation, respectively. After being infested by *T. urticae*, the amounts of chlorophyll from the sample and control leaves were 0.48 ± 0.02 and 0.42 ± 0.02 mg per gram fresh weight, respectively. The chlorophyll lost from the sample leaves was significantly lower than that from the control leaves. Between infestation states, $F_{1,32} = 50.083$ and $P < 0.0001$; between volatiles, $F_{1,32} = 4.579$ and $P = 0.0401$; and the interaction of the infestation and the volatiles, $F_{1,32} = 6.580$ and $P = 0.0152$ (ANOVA). Thus, after exposure to the *T. urticae*-induced volatiles, lima bean leaves become less suitable as a food resource to *T. urticae*.

The present analyses suggest the following mechanism of plant–plant interactions among lima bean plants that are mediated by airborne signals. In response to spider mites, a lima bean plant induces expression of at least six defence genes and emits volatiles accordingly. After plants are exposed to the terpenoids in the volatiles, a LOX gene is induced in neighbouring plants, followed by activation of the JA signalling pathway, and then expression of multiple defence genes that JA mediates. These plants can prepare defences against the spider mites in advance. The *T. urticae*-induced volatiles are specific and different in chemical composition from the wound-induced volatiles. Neighbouring plants can recognize such specificity and differences between leaf volatiles.

These findings on airborne information transferred among plants are important for understanding multitrophic interactions in ecosystems as well as for developing new methods for plant protection against herbivores. □

Methods

Bioassay and chemical application

Throughout our experiments, we detached a primary leaf with the petiole from intact lima bean plants (2 to 3 weeks old) and immediately transferred it into a vial (6 ml) containing

distilled water. We then placed about 100 *T. urticae* females on the detached leaf. Two infested leaves were placed in a plastic cup (10-cm diameter, 5.5-cm depth). To prevent the escape of *T. urticae* from the infested emitter leaves, we lined the inner wall of the cup with wet cotton wool. Artificially wounded leaves were also prepared by punching 35 holes (5-mm diameter) into each leaf. We placed the cups holding the infested or artificially wounded leaves in a lidded glass container (7 litre) together with four to six unfested receiver leaves set in vials containing water or chemical solution(s). The experimental setup was maintained at $25 \pm 2^\circ\text{C}$, under conditions of 50% to 70% relative humidity and 16 hours of light/8 hours of darkness for 1 or 3 days. We took care to ensure that the receiver leaves were never infested by *T. urticae*.

β -ocimene, DMNT and TMTT were synthesized in the laboratory. Linalool was obtained from Tokyo Chemical Industry. The chemicals were dissolved in pure dichloromethane ($\mu\text{g } \mu\text{l}^{-1}$). Each chemical (10 μl) was impregnated into a piece of cotton wool. After evaporation of the solvent, we enclosed the piece of cotton wool in a container (7 litre) together with two unfested receiver leaves for 3 or 24 h under the conditions described above. All chemicals, except β -ocimene, were more than 98% pure. β -ocimene was a mixture of *E* (about 70%) and *Z* (about 30%) isomers.

RT–PCR (polymer chain reaction after reverse transcription of RNA) analysis

Total RNA was extracted from two leaves by means of the acid guanidinium-phenol-chloroform method¹⁷. First-strand cDNA was synthesized from 1 μg of total RNA using a primer (5'-CTTGACCATCTATCTCTTC-3', 5'-GATGCGGTCTTGAACCTGC-3', 5'-GGACTCTTTTGATTCTCATC-3', 5'-TGACAATATCTCTATGACTG-3' or 5'-TCCTCCAAATGCTCAAGTC-3'), which corresponded, respectively, to nucleotides 336 to 355, 881 to 900, 864 to 883, 1972 to 1991, or 937 to 956 in the sequence of *Phaseolus vulgaris* acidic chitinase¹⁸, basic chitinase¹⁹, basic β -1,3-glucanase²⁰, pLOX³¹ or PAL cDNA²². The first-strand cDNA was amplified by adding Taq DNA polymerase (Takara, Japan) and a further primer (5'-AGCAACAACGTTAATGTTGC-3', 5'-CTCAGCGCCC TCATATCC AG-3', 5'-TATGCTCTTTTCACTTCACC-3', 5'-CTAGCAACAACAGGCAACT-3' or 5'-AGGCTGCTGCCATTATGGAG-3'), which corresponded, respectively, to nucleotides 105 to 124, 259 to 278, 582 to 601, 1260 to 1279, or 314 to 333 in the sequences of *P. vulgaris* acidic chitinase, basic chitinase, basic β -1,3-glucanase, LOX, or PAL cDNA in those reports. Synthesis consisted of 30, 25, 23, or 25 cycles (under optimum dynamic ranges before arriving at the plateau) of 95°C for 40 s, 55°C for 90 s and 72°C for 60 s for the amplifications of acidic chitinase, basic chitinase, basic β -1,3-glucanase, pLOX, and PAL cDNA, respectively. FPS cDNA was amplified with a pair of primers, 5'-CATGGATGAC TCTCACTC-3' and 5'-AGTCATCTGGACTTGAAG-3' which corresponded, respectively, to nucleotides 405 to 424 and 801 to 820 in the sequence of maize FPS cDNA²³ during 35 cycles (at 95°C for 40 s, 55°C for 90 s and 72°C for 1 min), after the first-strand cDNA was synthesized using a random nonamer primer. An equal amount of PCR products and 5 μg of total RNA were separated by electrophoresis in 1.5% agarose gels and detected by staining with ethidium bromide. No PCR products were amplified without reverse-transcription reactions, and the sequence analysis confirmed that each cDNA product was amplified from the transcribed product of lima bean gene.

Enzyme assay

We prepared enzyme extract by homogenization of a leaf with a grinding buffer containing 100 mM sodium phosphate (pH 7.0), 7% polyvinylpyrrolidone and 1% Triton X-100, followed by centrifugation. We measured the LOX activity²⁴ and protein concentrations²⁵ in the supernatant.

Analysis of volatile compounds

We collected the volatile compounds emitted from the two leaves in a lidded glass bottle (2 litre) using a glass desorption tube (3.0 mm i.d. $\times$ 160 mm length) packed with 100 mg of Tenax-TA resin (20/35 mesh; GL sciences, Japan) at a 100 ml min⁻¹ flow rate for 1 h. The absorbed compounds were eluted with 2 ml of diethyl ether, and then *n*-eicosane was immediately added to the eluate as the internal standard. After the eluate was concentrated with a stream of gaseous N_2 , it was injected into an injection port (250°C) of a gas chromatograph-mass spectrometer (GC-MS) (GC: Hewlett Packard 6890 with an HP-5MS capillary column with 0.25 mm i.d., 30 m length, 0.25 μm film thickness. MS: Hewlett Packard 5973 mass selective detector, 70 eV). The GC oven temperature was programmed to rise from 40°C (5 min. hold) to 280°C at $15^\circ\text{C min}^{-1}$.

Received 30 December 1999; accepted 30 May 2000.

1. Takabayashi, J. & Dicke, M. Plant-carnivore mutualism through herbivore-induced carnivore attractants. *Trends Plant Sci.* **1**, 109–113 (1996).
2. Dicke, M. *et al.* Isolation and identification of volatile kairomone that affects acarine predator-prey interactions. Involvement of host plant in its production. *J. Chem. Ecol.* **16**, 381–396 (1990).
3. Dicke, M. *et al.* Plant strategies of manipulating predator-prey interactions through allelochemicals: prospects for application in pest control. *J. Chem. Ecol.* **16**, 3091–3118 (1990).
4. Dicke, M., Gols, R., Ludeking, D. & Posthumus, M. A. Jasmonic acid and herbivory differentially induce carnivore-attracting plant volatiles in Lima bean plants. *J. Chem. Ecol.* **25**, 1907–1922 (1999).
5. De Moraes, C. M. *et al.* Herbivore-infested plants selectively attract parasitoids. *Nature* **393**, 570–573 (1998).
6. Bruin, J., Dicke, M. & Sabelis, M. W. Plants are better protected against spider-mites after exposure to volatiles from infested conspecifics. *Experientia* **48**, 525–529 (1992).
7. Ozawa, R. *et al.* Involvement of jasmonate- and salicylate-related signaling pathways for the production of specific herbivore-induced volatiles in plants. *Plant Cell Physiol.* **41**, 391–398 (2000).

8. Farmer, E. E. & Ryan, C. A. Interplant communication: Airborne methyl jasmonate induces synthesis of proteinase inhibitors in plant leaves. *Proc. Natl Acad. Sci. USA* **87**, 7713–7716 (1990).
9. McConn, M. et al. Jasmonate is essential for insect defence in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **94**, 5473–5477 (1997).
10. Brederode, F. T., Linthorst, H. J. M. & Bol, J. F. Differential induction of acquired resistance and PR gene expression in tobacco by virus infection, ethephon treatment, UV light and wounding. *Plant Mol. Biol.* **17**, 1117–1125 (1991).
11. Ryals, J. A. et al. Systemic acquired resistance. *Plant Cell* **8**, 1809–1819 (1996).
12. Bell, E., Creelman, R. A. & Mullet, J. E. A chloroplast lipoxygenase is required for wound-induced jasmonic acid accumulation in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **92**, 8675–8679 (1995).
13. Shulaev, V., Silverman, P. & Raskin, I. Airborne signaling by methyl salicylate in plant pathogen resistance. *Nature* **385**, 718–721 (1997).
14. Kahl, J. et al. Herbivore-induced ethylene suppresses a direct defence but not a putative indirect defense against an adapted herbivore. *Planta* **210**, 336–342 (2000).
15. Bate, N. J. & Rothstein, S. J. C₆-volatiles derived from the lipoxygenase pathway induce a subset of defence-related genes. *Plant J.* **16**, 561–569 (1998).
16. Mackinney, G. Absorption of light by chlorophyll solutions. *J. Biol. Chem.* **140**, 315–322 (1941).
17. Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159 (1987).
18. Margis-Pinheiro, M. et al. Isolation of a complementary DNA encoding the bean PR4 chitinase: an acidic enzyme with an amino-terminus cysteine-rich domain. *Plant Mol. Biol.* **17**, 243–253 (1991).
19. Broglie, K. E., Gaynor, J. J. & Broglie, R. M. Ethylene-regulated gene expression: Molecular cloning of the genes encoding an endochitinase from *Phaseolus vulgaris*. *Proc. Natl Acad. Sci. USA* **83**, 6820–6824 (1986).
20. Edington, B. V., Lamb, C. J. & Dixon, R. A. cDNA cloning and characterization of a putative 1,3-β-D-glucanase transcript induced by fungal elicitor in bean cell suspension cultures. *Plant Mol. Biol.* **16**, 81–94 (1991).
21. Meier, B. M., Shaw, N. & Shusarenko, A. J. Spatial and temporal accumulation of defense gene transcripts in bean (*Phaseolus vulgaris*) leaves in relation to bacteria-induced hypersensitive cell death. *Mol. Plant Microbe Interact.* **6**, 453–466 (1993).
22. Edwards, K. et al. Rapid transient induction of phenylalanine ammonia-lyase mRNA in elicitor-treated bean cells. *Proc. Natl Acad. Sci. USA* **82**, 6731–6735 (1985).
23. Li, C. P. & Larkins, B. A. Identification of a maize endosperm-specific cDNA encoding farnesyl pyrophosphate synthetase. *Gene* **171**, 193–196 (1996).
24. Hildebrand, D. F. et al. Peroxidative responses of leaves in two soybean genotypes injured by twospotted spider mites (Acari: Tetranychidae). *J. Econ. Entomol.* **79**, 1459–1465 (1986).
25. Bradford, M. M. A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).

Acknowledgements

We thank J. Horiuchi for assistance in experiments, and M. Dicke, M. W. Sabelis, J. Bruin and M. Kant for their comments on the manuscript. This study was partly supported by the Program for Promotion of Basic Research Activities for Innovative Biosciences (Bio-oriented Technology Research Advancement Institution) and by a Research for the Future program (JSPS-RFTF96L006703).

Correspondence and requests for materials should be addressed to J.T.
(e-mail: junji@kais.kyoto-u.ac.jp).

Genetic and epigenetic mechanisms contribute to motor neuron pathfinding

Kamal Sharma*, Ann E. Leonard, Karen Lettieri & Samuel L. Pfaff

Gene Expression Laboratory, The Salk Institute for Biological Studies, La Jolla, California 92037, USA

Many lines of evidence indicate that genetically distinct subtypes of motor neurons are specified during development¹, with each type having characteristic properties of axon guidance and cell-body migration². Motor neuron subtypes express unique combinations of LIM-type homeodomain factors that may act as intrinsic genetic regulators of the cytoskeletal events that mediate cell migration, axon navigation or both^{3–7}. Although experimentally displaced motor neurons can pioneer new routes to their targets^{8–11}, in many cases the axons of motor neurons in complete

isolation from their normal territories passively follow stereotypical pathways dictated by the environment^{12–16}. To investigate the nonspecific versus genetically controlled regulation of motor connectivity we forced all motor neurons to express ectopically a LIM gene combination appropriate for the subgroup that innervates axial muscles. Here we show that this genetic alteration is sufficient to convert the cell body settling pattern, gene-expression profile and axonal projections of all motor neurons to that of the axial subclass. Nevertheless, elevated occupancy of the axial pathway can override their genetic program, causing some axons to project to alternative targets.

At the birth of motor neurons the genes *Isl1*, *Lhx3* (*Lim3*) and the redundant gene *Lhx4* (*Gsh4*) act together to direct axons ventrally from the neural tube¹⁷. This early combinatorial pattern of LIM gene expression changes rapidly as motor axons emerge ventrally from the neural tube, so that each ventral-exiting motor column subtype expresses a distinct combination of factors (Fig. 1). To examine whether LIM homeodomain transcription factors define motor column identity we altered their combinatorial pattern of expression in developing motor neurons. We sought to convert all motor neurons to an axial-innervating (medial half of median motor column, MMCm; Fig. 1) identity by ectopically expressing *Lhx3*, a LIM homeodomain factor specific to the MMCm subtype. We used the promoter of the motor-neuron-specific gene *HB9* (refs 18, 19) to drive expression of *Lhx3* (Fig. 2a). In two control mouse lines we expressed nuclear-localized β-galactosidase (nLacZ) or the LMC-motor neuron LIM homeodomain factor *Lhx1* (*Lim-1*) using the same strategy (Fig. 2a). We used a Cre-removable lox-neomycin-lox (LNL) block on transgene translation to avoid the expected lethality of severe motor defects. One copy of each

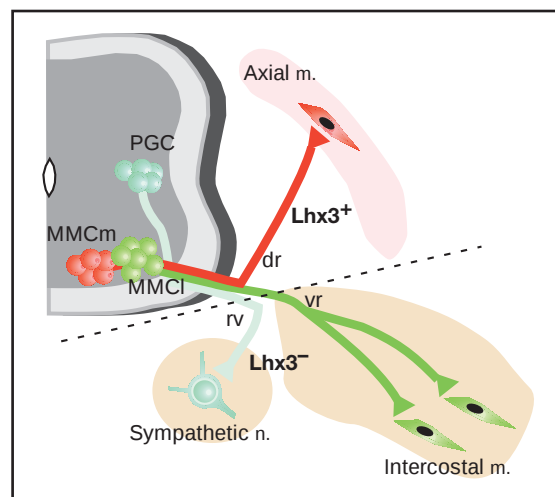


Figure 1 Summary of thoracic motor columns. Motor columns express unique combinations of LIM homeodomain transcription factors which emerge from earlier patterns of LIM factors at motor neuron birth¹⁷. MMCm cells (red) express *Isl1*, *Isl2*, *Lhx3*, *Lhx4* and *HB9*. *Lhx3* and *Lhx4* have redundant activities¹⁷, therefore only *Lhx3* is shown. MMCm motor neurons project axons to axial muscles through the dorsal root (dr) pathway at thoracic levels. At lower cervical levels the axial musculature is more complex and MMCm motor neurons use several branches to innervate these targets. MMCl cells (green) project axons into the ventral root (vr) and express *Isl1*, *Isl2* and *HB9*. PGC cells (blue) project axons into the ramus visceralis (rv) and express *Isl1* and *HB9*. At limb levels (lower cervical and lumbar) LMC motor neurons (*Isl1*, *Isl2*, *Lhx1*, *HB9*) occupy a position similar to the PGC in the spinal cord and project axons along a ventral root-like pathway to plexi at the base of the limb. The spinal cord at limb levels does not contain PGC or MMCl cells. MMCm, medial half of median motor column; MMCl, lateral half of median motor column; PGC, preganglionic motor column; LMC, lateral motor column; m, muscle; n, neuron.

* Present address: Department of Neurobiology, University of Chicago, Chicago, Illinois 60637, USA.

transgene was inserted into the 3' noncoding region of the *Isl2* locus, creating the mouse lines TgH *Isl2*^{HB9:LNL:nLacZ}, TgH *Isl2*^{HB9:LNL:Lhx1} and TgH *Isl2*^{HB9:LNL:Lhx3} (Fig. 2a). This knock-in strategy eliminated unpredictable enhancer/suppressor effects from random integration sites and provided a defined substrate for DNA recombination by Cre. TgH animals were mated to the Protamine promoter-Cre line TgN^{Pro:Cre} (ref. 20) (Fig. 2).

As desired, expression of nLacZ, Lhx1 and Lhx3 was detected in HB9⁺ motor neurons in recombined (LNL converted to L) F2 embryos (TgH *Isl2*^{HB9:LNL:nLacZ}, TgH *Isl2*^{HB9:LNL:Lhx1} and TgH *Isl2*^{HB9:LNL:Lhx3}), respectively (Fig. 2c, e, g). We abbreviate the genotypes of these embryos to TgH *nLacZ*^{ON}, TgH *Lhx1*^{ON} and TgH *Lhx3*^{ON}. Transgene expression was detected in early postmitotic motor neurons before the establishment of definitive motor columns, at levels that were comparable to endogenous LIM genes (Fig. 2c, e, g). As predicted, non-recombined F2 embryos did not show detectable transgene expression (Fig. 2b, d, f); therefore we abbreviate their genotypes to TgH *nLacZ*^{OFF}, TgH *Lhx1*^{OFF} and TgH *Lhx3*^{OFF}.

Embryos with either recombined (ON) or non-recombined (OFF) alleles of TgH *nLacZ*, TgH *Lhx1* and TgH *Lhx3* developed normally. Postnatally, both TgH *nLacZ*^{ON} and TgH *nLacZ*^{OFF} animals were normal and found at their expected frequencies. Whereas TgH *Lhx3*^{OFF} animals were viable postnatally, TgH

Lhx3^{ON} animals died at birth from apparent respiratory failure. The dominant lethal phenotype in TgH *Lhx3*^{ON} postnatal mice was consistent with a motor defect, although normal numbers of motor neurons were generated in these animals (see below).

We expected that the ectopic expression of Lhx3 would prevent the diversification of motor neuron subtypes and instead drive cells toward an MMCm identity (Fig. 1). First we examined the expression of *Lhx4* in TgH *Lhx3*^{ON} embryos, because this MMCm-specific gene is not a direct transcriptional target of Lhx3 (ref. 17). Unlike TgH *Lhx3*^{OFF} embryos, in which Lhx4 labels only the MMCm fraction of motor neurons, TgH *Lhx3*^{ON} embryos show Lhx4 expression in all motor neurons (Fig. 3a, b). Embryonic stage (E) 15.5 TgH *Lhx3*^{ON} embryos contain 40 ± 8 Lhx4⁺ cells per 12- μ m thoracic section, compared with 20 ± 2 cells in TgH *Lhx3*^{OFF} embryos.

If motor neurons truly acquire an MMCm identity in TgH *Lhx3*^{ON} embryos, markers of other motor neuron subtypes are expected to be downregulated. Motor neurons in the preganglionic motor column (PGC) express NADPH diaphorase²¹, an enzyme in the nitric oxide pathway. No diaphorase activity was detected in TgH *Lhx3*^{ON} embryos (Fig. 3c, d). Likewise, TgH *Lhx3*^{ON} embryos expressed neither the Lhx1 marker of the lateral half of the lateral motor column (LMC1)³ (not shown) nor Ets transcription factor *Er81*, which defines individual motor pools in the LMC²² (Fig. 3e, f). Together, our results provide strong evidence that motor neurons in TgH *Lhx3*^{ON} embryos homogeneously express molecular attributes of MMCm cells.

To control for the specificity of Lhx3 activity, we examined motor

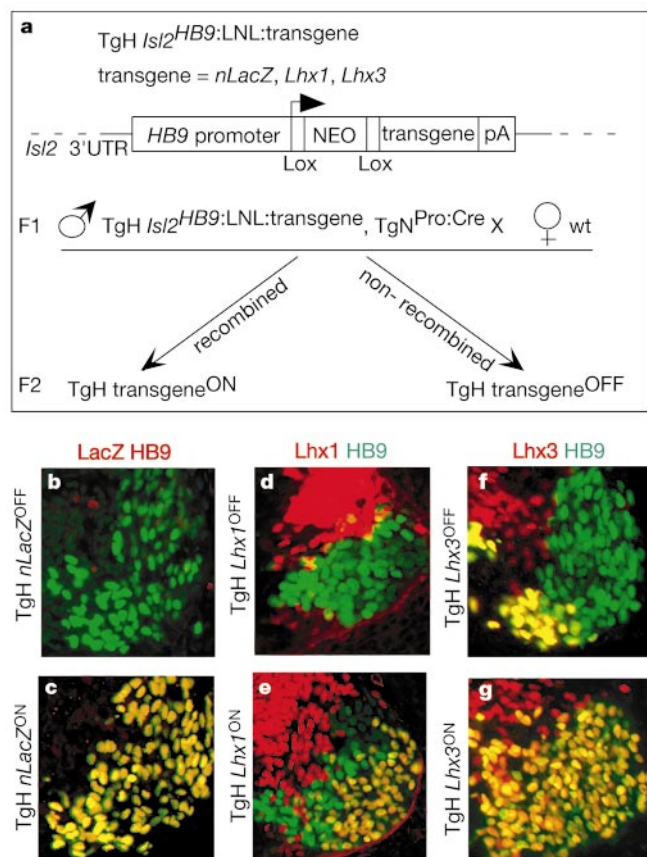


Figure 2 Method for ectopic gene expression in embryonic motor neurons. **a**, Schematic of HB9 promoter construct integrated into the 3'-untranslated region of *Isl2* for expression of *nLacZ*, *Lhx1* and *Lhx3* transgenes. Male germline expression of Cre is used to delete the lox-neomycin-lox (LNL) cassette, permitting transgene translation. **b-g**, E11.5 embryos immunostained to detect *nLacZ*, *Lhx1*, *Lhx3* and HB9 in the right ventral quadrant of the spinal cord. **b, d, f**, Non-recombined embryos (OFF) showed no transgene expression. *Lhx1* (**d**) and *Lhx3* (**f**) are normally expressed in interneurons (red) and subclasses of motor neurons³. **c, e, g**, Recombined embryos (ON) expressed transgenes ectopically in HB9⁺ motor neurons as they were born. $n > 5$ embryos for each genotype.

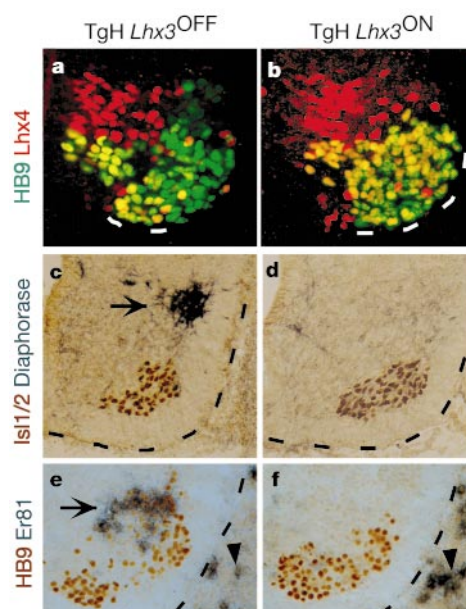


Figure 3 Ectopic Lhx3 induces MMCm markers in motor neurons. **a, b**, Thoracic immunocytochemical staining of E11.5 embryos with Lhx4 and HB9. **a**, Lhx4 is expressed in interneurons (red) and HB9⁺ MMCm motor neurons (yellow). Non-MMCm motor neurons do not express Lhx4 (green). **b**, TgH *Lhx3*^{ON} embryos display Lhx4 expression in all HB9⁺ motor neurons, indicative of MMCm cell identity. **c, d**, NADPH diaphorase immunohistochemistry (black) and Isl1/2 immunocytochemistry (brown) on E13.5 thoracic cords. **c**, PGC cells in the intermediolateral spinal cord are double labelled in TgH *Lhx3*^{OFF} embryos (arrow). **d**, No diaphorase activity or intermediolateral Isl1⁺ cells were detected in TgH *Lhx3*^{ON} embryos. **e, f**, *Er81* *in-situ* hybridization (black) and HB9 immunostaining (brown) on E13.5 lumbar cords. **e**, LMC motor pools are double labelled in TgH *Lhx3*^{OFF} embryos (arrow). **f**, No *Er81* expression is detected in motor neurons of TgH *Lhx3*^{ON} embryos, although it is detected in dorsal root ganglion cells (**e, f**, arrowheads). $n > 20$ sections, $n > 3$ embryos for each sample.

neuron development in TgH *Lhx1*^{ON} embryos. TgH *Lhx1*^{ON} embryos did not display any evidence of ectopic MMCm generation, on the basis of molecular markers, cell migration or axon guidance (not shown). Likewise, motor neuron subtypes form normally in TgH *nLacZ*^{OFF} and TgH *nLacZ*^{ON} embryos (not shown).

The analysis of TgH *Lhx3*^{ON} embryos described above revealed changes in the settling pattern of motor neuron cell bodies. MMCm neurons settle in a ventromedial position, whereas PGC and LMC motor neurons migrate from the ventral horn to settle along the intermediolateral edge of the spinal cord (Fig. 1). In E15.5 TgH *Lhx3*^{ON} embryos, no (0 ± 0) Isl1⁺ cells in the intermediolateral position of PGC motor neurons were detected at thoracic levels, whereas controls had 12 ± 2 intermediolateral PGC cells per 12-μm section (Fig. 3c, d). This was examined more precisely using retrograde labelling with rhodamine-dextran to identify specific intermediolateral cell types. Injections in the sympathetic ganglia labelled fewer PGC cells in TgH *Lhx3*^{ON} embryos than in TgH *Lhx3*^{OFF} embryos, reflecting a reduction in the number of motor neurons projecting to this target (see below). Motor neurons that

had projected to the sympathetic ganglia in TgH *Lhx3*^{ON} embryos were located in the ventral horn rather than the normal intermediolateral position (Fig. 4a–d).

LMC motor neurons also failed to migrate from the ventral spinal cord to their normal intermediolateral position. Double-labelling was performed to define the ventromedial motor neurons of the MMCm with FITC-dextran injections into axial muscles, and intermediolateral cells of the LMC were labelled with rhodamine-dextran injections into the limb bud. In TgH *Lhx3*^{OFF} embryos, MMCm and LMC motor neurons were clearly segregated into distinct columns (Fig. 4e, g). In contrast, cells backlabelled from the limb bud of TgH *Lhx3*^{ON} embryos were intermingled with the MMCm (Fig. 4f, h). Rather than forming distinct cell clusters, motor neurons settle ventrally as a group in *Lhx3*^{ON} embryos. These data are consistent with the notion that *Lhx3* expression confers the MMCm identity.

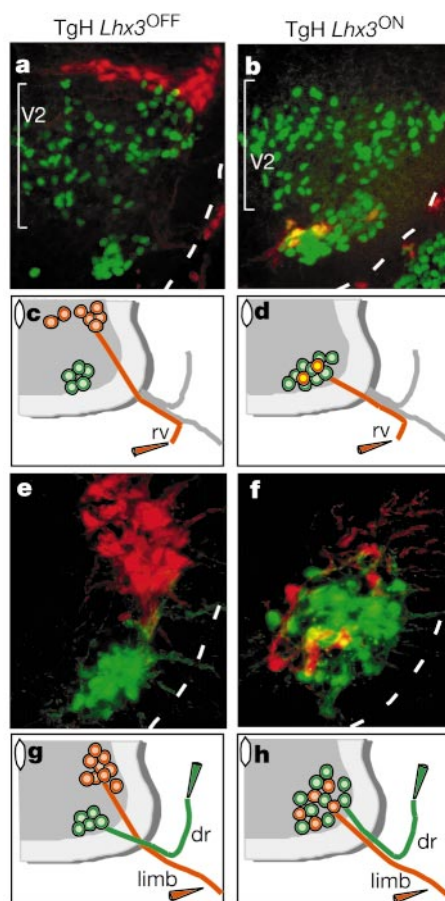


Figure 4 Analysis of motor neuron settling patterns. **a–d**, E13.5 thoracic spinal cords retrogradely labelled with rhodamine-dextran from the sympathetic ganglia and immunostained to detect *Lhx3* (green). **a**, PGC cells (red) are dorsal to a class of *Lhx3*⁺ ventral interneurons (V2, bracket). **b**, Cells that project to the sympathetic ganglia (red) are reduced in number and are inappropriately located in the ventral neural tube of TgH *Lhx3*^{ON} embryos. **c, d**, Schematic of ramus visceralis (rv) fills in **a, b**, respectively. **e–h**, E13.5 cervical spinal cords retrogradely labelled from the limb with rhodamine-dextran and axial muscles with FITC-dextran. **e**, LMC motor neurons (red) are dorsal to MMCm motor neurons (green). **f**, Limb-projecting motor neurons (red) are reduced in number and intermingled with axial-projecting motor neurons (green) in TgH *Lhx3*^{ON} embryos. **g, h**, Schematic of limb and dorsal ramus fills in **e, f**, respectively. *n* > 20 embryos.

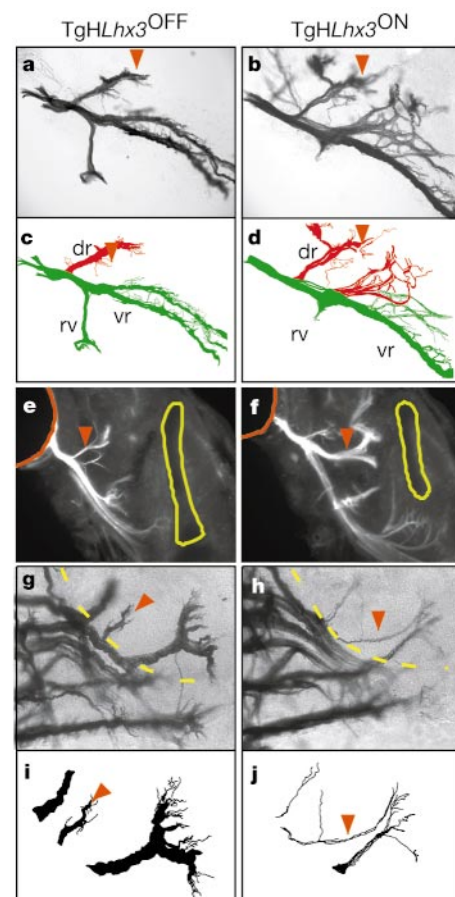


Figure 5 Orthograde fills reveal increased axial projections in TgH *Lhx3*^{ON} embryos. **a–j**, Orthograde Dil labelling of motor axons in E13.5 embryos. **a, b**, At thoracic levels three branches of axons are labelled, the dorsal ramus (dr, red arrowhead), the ventral ramus (vr) which splits distally, and the ramus visceralis (rv). Axial muscle projections increase in TgH *Lhx3*^{ON} embryos (red arrowhead), whereas the ramus visceralis is greatly reduced. **c, d**, Trace drawing of axial projections (red) and non-axial projections (green) in **a, b**, respectively. Many axons enter the ventral ramus but then branch back to the axial musculature in TgH *Lhx3*^{ON} embryos. **e**, Several nerve branches innervate the axial musculature at cervical levels. **f**, In TgH *Lhx3*^{ON} embryos the axial nerves are enlarged (compare arrowheads in **e, f**). Yellow and red correspond to the outline of the scapula and spinal cord, respectively. **g**, Projections distal to the brachial plexus have been labelled in the limb bud (yellow dotted line separates plexus from limb nerves). **h**, Fewer axons enter the limb of TgH *Lhx3*^{ON} embryos. Most axon pathways are labelled, though a few minor misprojections are detected (arrowhead). **i, j**, Trace drawing of muscle nerve branches in **g, h**, respectively. *n* > 10 embryos for each genotype.

If LIM factors regulate axonal projections of motor column subtypes, the ectopic expression of *Lhx3* is predicted to redirect motor axons towards axial muscles (Fig. 1). The double labelling of axial- and limb-innervating motor neurons in TgH *Lhx3*^{ON} embryos showed that axial innervation was increased, with a concomitant reduction in limb innervation (compare Fig. 4e and f). We used orthograde DiI fills of motor axons to examine axonal projections directly. At thoracic levels motor neurons innervate axial muscles, intercostal muscles and sympathetic ganglia. In TgH *Lhx3*^{ON} embryos, axons arborize extensively within axial muscles, but the ramus visceralis branch is nearly absent (Fig. 5a–d). At cervical spinal levels, nerves to axial muscles are clearly enlarged (Fig. 5e, f). Many axons that first bypassed axial nerve pathways such as the dorsal ramus redirected their growth distally toward axial muscles, whereas still others simply followed the normal ventral ramus pathway (Fig. 5d).

Next, we examined the innervation of the limb bud using orthograde DiI fills. The number of axons labelled in the limb of TgH *Lhx3*^{ON} embryos was greatly reduced compared with controls (Fig. 5g–j). This appeared to be due to an increase in the projections to axial targets combined with the apparent stopping of *Lhx3*⁺ axons at the base of the limb. The axons that did enter the limb were positioned normally, but a few subtle pathfinding defects were observed in TgH *Lhx3*^{ON} embryos. Although we predicted that *Lhx3* expression would direct all motor neurons towards axial targets we found that the penetrance of this phenotype was incomplete. Once *Lhx3*⁺ axons bypassed axial targets in TgH *Lhx3*^{ON} embryos they generally followed the nerve pathways used by *Lhx3*[−] motor neurons.

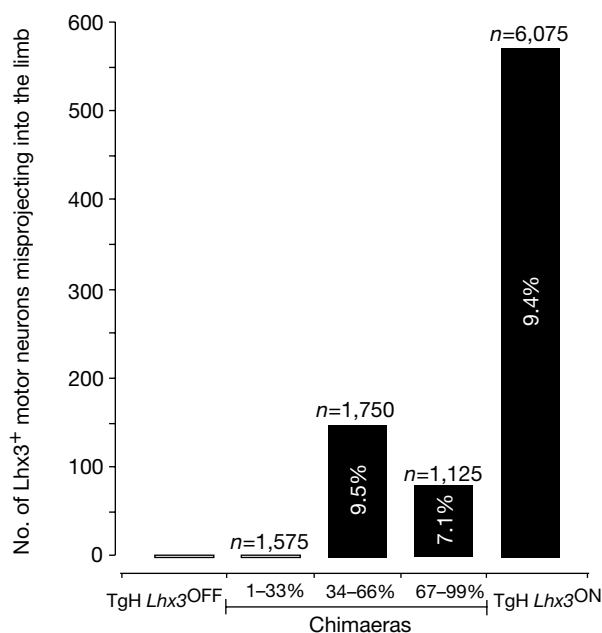


Figure 6 MMCm cell number influences pathfinding. Forelimb projections of *Lhx3*⁺ cells were identified by backlabelling in chimaeric TgH *Lhx3*^{ON} ↔ wild-type embryos. Misprojections were plotted on the y axis, defined as *Lhx3*⁺ cells that backlabelled from the forelimb. The total number of *Lhx3*⁺ motor neurons was monitored in hemi-segments C5–6. Data from embryos with similar *Lhx3* numbers was binned (13 embryos for 1–33% chimaeras, 4 embryos for 34–66% chimaeras, and 1 embryo for TgH *Lhx3*^{OFF}, TgH *Lhx3*^{ON}, and 67–99% chimaeras). The *n* value represents the number of ectopic *Lhx3*⁺ motor neurons examined in each bin, and the percentage shown in the columns corresponds to the detected pathfinding error frequency. *Lhx3* is sufficient to direct axons away from the limb in embryos with a low degree of chimaerism (1–33%), but insufficient when many MMCm (*Lhx3*⁺) motor neurons are present.

The presence of non-axial nerves in TgH *Lhx3*^{ON} embryos indicated either that *Lhx3* was insufficient to regulate genes for axon navigation to axial muscles or that some other feature of *Lhx3*^{ON} embryos limited axial muscle innervation. To distinguish between these possibilities we assessed motor axon projections in embryos with a small increase in *Lhx3*⁺ cell numbers by creating aggregation chimaeras between LacZ-marked TgH *Lhx3*^{ON} embryos and wild-type embryos. LacZ and *Lhx3* immunostaining were used to monitor the degree of chimaerism at each spinal cord segment. E13.5 TgH *Lhx3*^{ON} embryos average 111 *Lhx3*⁺ motor neurons per 12-μm section at cervical (C) levels 5/6, whereas wild-type embryos have 30 *Lhx3*⁺ cells per section. Backlabels from the limbs of wild-type embryos fill 0 ± 0 *Lhx3*⁺ cells, whereas fills of TgH *Lhx3*^{ON} embryos consistently label about 10 *Lhx3*⁺ cells per 12-μm section (9.4% of the total motor neurons, Fig. 6). We define these as ‘misprojections’ because of the discrepancy between LIM code and axon target. Thirteen chimaeric embryos were generated with 32–56 *Lhx3*⁺ cells per section (corresponding to a 1–33% degree of chimaerism). Of the 1,575 ectopic motor neurons examined in these embryos, no misprojections were detected (Fig. 6). In contrast, embryos with a higher percentage of *Lhx3*⁺ motor neurons exhibited many misprojections among a similar total number of ectopic *Lhx3*⁺ cells (Fig. 6). These data show that *Lhx3* is sufficient to direct axons away from non-axial muscles in the 1–33% chimaeric embryos.

If *Lhx3* is sufficient to regulate projections into the dorsal ramus, why do TgH *Lhx3*^{ON} embryos misproject axons into the ventral ramus pathway to limb and intercostal muscles? One possibility is that the reduced innervation of the limb results in a change in the pattern of guidance cue expression that may begin to attract *Lhx3*⁺ axons. This seems unlikely, however, given that partial ablation of the spinal cord results in limb denervation but not ectopic innervation²³. As *Lhx3*⁺ motor neurons backlabelled from the limb lacked the discrete cellular organization found in wild-type spinal cords (Fig. 4f), it appeared that random cells were misguided into the limb. Therefore we considered an alternative possibility in which the pathfinding of *Lhx3*⁺ motor neurons is influenced by axonal competition analogous to that observed for retinal–tectal projections²⁴. In contrast to wild-type and low level chimaeras, embryos with more than 56 *Lhx3*⁺ cells per section (>33% chimaerism) exhibited a consistent frequency of misprojections in the range 7.1%–9.5% (Fig. 6). Because the frequency of misprojections is relatively constant over a range of ectopic *Lhx3*⁺ motor neurons, it is likely that the number of axial projections increases as *Lhx3*⁺ cell numbers increase. The misprojections we observe in TgH *Lhx3*^{ON} embryos appear to result from the shunting of axons from heavily occupied axial targets or from inappropriate axons entering depleted pathways. In either case these findings reveal an epigenetic interaction that can override the intrinsic activity of LIM genes in these cells.

Together, these results show that LIM factors control the development of topological order among motor neuron subtypes by coordinately regulating both axon guidance and cell migration. A small number of LIM homeodomain factors appear to contribute to a large diversification of motor neuron subtypes by acting in a combinatorial manner and at several distinct phases of motor neuron development. As *Lhx3* is initially expressed in all motor neurons that send axons ventrally from the neural tube and then becomes restricted to the MMCm subtype¹⁷, our results show that a critical step for the specification of non-MMCm cells is the down-regulation of *Lhx3*. The choice between ISNb and ISNd axon projections in *Drosophila* is also regulated by the combinatorial activity of *islet* and *lim3* (*Lhx3/4* homologue)²⁵.

The penetrance of the axon rerouting phenotype induced by *Lhx3* expression is incomplete in both *Drosophila* and mouse. Using chimaeras, we show that axial pathway selection is strongly influenced when the axonal occupancy exceeds a threshold created by more than 56 *Lhx3*⁺ motor neurons per section. Counterintuitively,

once axons of $Lhx3^+$ motor neurons have bypassed axial muscle targets they continue to grow along $Lhx3^-$ pathways, lending further support to the notion that the passive deployment activity of motor axons is quite robust^{12–16}. Indeed, this passive activity combined with epigenetic interactions raises the possibility that knockouts of guidance receptors could appear to have all nerves in place, despite an erosion of specificity for targets. An ability to extend axons passively seems incongruous with the highly specific pathfinding observed for motor neurons². Our findings suggest an additional mechanism that may cooperate with guidance receptors to establish specific motor connections. In principle, the occupancy of a pathway could increase the relative barrier for inappropriate motor axons, thus magnifying the actions of guidance receptors and enhancing the fidelity of pathfinding. □

Methods

A 9-kilobase fragment of the upstream region of *HB9* (ref. 19) was linked to complementary DNAs encoding *nLacZ*, *Lhx1* or *Lhx3*. At the start methionine a lox–neomycin–lox cassette was inserted such that translation could occur following Cre-mediated DNA recombination at the lox sites. Promoter constructs were targeted by homologous recombination into the 3′-untranslated region of *Isl2*.

To generate aggregation chimaeras *TgH Isl2^{HB9:LN:Lhx3}/TgN^{ProCre}* F1 males were crossed to superovulated ROSA26 (ref. 26) females. Morula stage embryos were flushed, stripped of the Zona and aggregated to wild-type morula embryos, transferred into pseudopregnant females and recovered at E13.5 for analysis.

Immunocytochemistry, *in situ* hybridization and neuronal labelling were performed as described^{3,17,19,27}.

Received 22 February; accepted 22 May 2000.

1. Tanabe, Y. & Jessell, T. M. Diversity and pattern in the developing spinal cord. *Science* **274**, 1115–1122 (1996).
2. Landmesser, L. The development of motor projection patterns in the chick hind limb. *J. Physiol. (Lond.)* **284**, 391–414 (1978).
3. Tsuchida, T. *et al.* Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes. *Cell* **79**, 957–970 (1994).
4. Varela-Echavarría, A., Pfaff, S. L. & Guthrie, S. Differential expression of LIM homeobox genes among motor neuron subpopulations in the developing chick brain stem. *Mol. Cell. Neurosci.* **8**, 242–257 (1996).
5. Appel, B. *et al.* Motoneuron fate specification revealed by patterned LIM homeobox gene expression in embryonic zebrafish. *Development* **121**, 4117–4125 (1995).
6. Jurata, L. W., Thomas, J. B. & Pfaff, S. L. Transcriptional mechanisms in the development of motor control. *Curr. Opin. Neurobiol.* **10**, 72–79 (2000).
7. Pfaff, S. & Kintner, C. Neuronal diversification: development of motor neuron subtypes. *Curr. Opin. Neurobiol.* **8**, 27–36 (1998).
8. Bell, E., Wingate, R. J. & Lumsden, A. Homeotic transformation of rhombomere identity after localized *Hoxb1* misexpression. *Science* **284**, 2168–2171 (1999).
9. Ferguson, B. A. Development of motor innervation of the chick following dorsal-ventral limb bud rotations. *J. Neurosci.* **3**, 1760–1772 (1983).
10. Lance-Jones, C. & Landmesser, L. Motoneurone projection patterns in the chick hind limb following partial reversals of the spinal cord. *J. Physiol.* **302**, 581–602 (1980).
11. Ferns, M. J. & Hollyday, M. Motor innervation of dorsoventrally reversed wings in chick/quail chimeric embryos. *J. Neurosci.* **13**, 2463–2476 (1993).
12. Summerbell, D. & Stirling, R. V. The innervation of dorsoventrally reversed chick wings: evidence that motor axons do not actively seek out their appropriate targets. *J. Embryol. Exp. Morphol.* **61**, 233–247 (1981).
13. Straznick, C. The patterns of innervation and movements of ectopic hindlimb supplied by brachial spinal cord segments in the chick. *Anat. Embryol. (Berl.)* **167**, 247–262 (1983).
14. Lance-Jones, C. C. Motoneurone projection patterns in chick embryonic limbs with a double complement of dorsal thigh musculature. *Dev. Biol.* **116**, 387–406 (1986).
15. Whitelaw, V. & Hollyday, M. Position-dependent motor innervation of the chick hindlimb following serial and parallel duplications of limb segments. *J. Neurosci.* **3**, 1216–1225 (1983).
16. O'Brien, M. K., Landmesser, L. & Oppenheim, R. W. Development and survival of thoracic motoneurons and hindlimb musculature following transplantation of the thoracic neural tube to the lumbar region in the chick embryo: functional aspects. *J. Neurobiol.* **21**, 341–355 (1990).
17. Sharma, K. *et al.* LIM homeodomain factors *Lhx3* and *Lhx4* assign subtype identities for motor neurons. *Cell* **95**, 817–828 (1998).
18. Arber, S. *et al.* Requirement for the homeobox gene *Hb9* in the consolidation of motor neuron identity. *Neuron* **23**, 659–674 (1999).
19. Thaler, J. *et al.* Active suppression of interneuron programs within developing motor neurons revealed by analysis of homeodomain factor *HB9*. *Neuron* **23**, 675–687 (1999).
20. O'Gorman, S., Dagenais, N. A., Qian, M. & Marchuk, Y. Protamine-Cre recombinase transgenes efficiently recombine target sequences in the male germ line of mice, but not in embryonic stem cells. *Proc. Natl Acad. Sci. USA* **94**, 14602–14607 (1997).
21. Wetts, R. & Vaughn, J. E. Choline acetyltransferase and NADPH diaphorase are co-expressed in rat spinal cord neurons. *Neuroscience* **63**, 1117–1124 (1994).
22. Lin, J. H. *et al.* Functionally related motor neuron pool and muscle sensory afferent subtypes defined by coordinate ETS gene expression. *Cell* **95**, 393–407 (1998).
23. Lance-Jones, C. & Landmesser, L. Motoneurone projection patterns in embryonic chick limbs following partial deletions of the spinal cord. *J. Physiol. (Lond.)* **302**, 559–580 (1980).
24. Goodhill, G. J. & Richards, L. J. Retinotectal maps: molecules, models and misplaced data. *Trends Neurosci.* **22**, 529–534 (1999).

25. Thor, S., Andersson, S. G., Tomlinson, A. & Thomas, J. B. A LIM-homeodomain combinatorial code for motor-neuron pathway selection. *Nature* **397**, 76–80 (1999).
26. Friedrich, G. & Soriano, P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* **5**, 1513–1523 (1991).
27. Pfaff, S. L., Mendelsohn, M., Stewart, C. L., Edlund, T. & Jessell, T. M. Requirement for LIM homeobox gene *Isl1* in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. *Cell* **84**, 309–320 (1996).

Acknowledgements

We thank A. Joyner and A. Auerbach for the training to generate chimaeric embryos; M. McLean for help with immunocytochemistry; S. O'Gorman for Cre mice and the lox–neo–lox cassette; B. Hogan for *Lhx1*; H. Westphal for *Lhx3*; L. Jurata, C. Lance-Jones, G. Lemke and J. Thomas; and J. Thaler for insightful comments on the manuscript. This work derives from initial studies on LIM genes while S.L.P. was a fellow with T. Jessell. The Human Frontiers Science Program provided postdoctoral support for K. S., and the National Institutes of Health provided training support for A.E.L. This research was funded by the National Institutes of Health. S.L.P. is a Basil O'Connor, McKnight, Pew, and Alfred P. Sloan Scholar.

Correspondence and requests for materials should be addressed to S.L.P. (e-mail: Pfaff@Salk.edu).

An intrinsic but cell-nonautonomous defect in GATA-1-overexpressing mouse erythroid cells

David Whyatt*, Fokke Lindeboom†, Alar Karis†, Rita Ferreira†, Eric Milot†, Rudi Hendriks†, Marella de Bruijn†, An Langeveld†, Joost Gribnau†, Frank Grosveld† & Sjaak Philipsen†

* Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands

† Erasmus University, Department of Cell Biology and Genetics, Medical Genetics Centre, PO Box 1738, 3000 DR Rotterdam, The Netherlands

GATA-1 is a tissue-specific transcription factor that is essential for the production of red blood cells^{1,2}. Here we show that overexpression of GATA-1 in erythroid cells inhibits their differentiation, leading to a lethal anaemia. Using chromosome-X-inactivation of a GATA-1 transgene and chimaeric animals, we show that this defect is intrinsic to erythroid cells, but nevertheless cell nonautonomous. Usually, cell nonautonomy is thought to reflect aberrant gene function in cells other than those that exhibit the phenotype³. On the basis of our data, we propose an alternative mechanism in which a signal originating from wild-type erythroid cells restores normal differentiation to cells overexpressing GATA-1 *in vivo*. The existence of such a signalling mechanism indicates that previous interpretations of cell-nonautonomous defects may be erroneous in some cases and may in fact assign gene function to incorrect cell types.

GATA-1 expression is required at the relatively immature proerythroblast stage^{1,2}, as GATA-1 null proerythroblasts undergo apoptosis⁴ and reduced GATA-1 levels inhibit proerythroblast differentiation⁵. However, the role of GATA-1 later in erythroid differentiation remains obscure. Murine erythroleukaemia (MEL) cells overexpressing GATA-1 under the control of the erythroid-specific human β -globin locus control region linked to the β -globin promoter (construct PEV3–GATA-1) fail to activate the expression of differentiation markers in response to the chemical inducer dimethyl sulphoxide (DMSO) and do not undergo differentiation-associated proliferative arrest⁶. Embryonic stem (ES) cell clones overexpressing GATA-1 also generate erythroid colonies that are inhibited in terminal differentiation⁶. Furthermore, overexpression of an inducible GATA-1 fusion protein (GATA-1–LBD, containing the tamoxifen-inducible ligand-binding domain of the

oestrogen receptor) also inhibits erythroid differentiation (R.F. and D.W., unpublished data). These results may explain our failure to produce a transgenic line of mice by conventional microinjection of PEV3–GATA-1, because loss of erythroid differentiation would be lethal *in vivo*.

We therefore exploited the process of X-inactivation⁷, as an X-linked GATA-1 transgene should be transcriptionally active in 50% of female erythroid precursors. In males, X-inactivation does not occur, and all erythroid cells should overexpress GATA-1. Mice can survive to term when 50% of erythroid precursors fail to differentiate normally, for example in females with one disrupted GATA-1 allele (which is X-linked); however, roughly 30% of GATA-1 null heterozygous females die from severe anaemia by 15.5 days post coitus (d.p.c.). In survivors, anaemia is transient and recovery is thought to be due to *in vivo* selection of normal progenitors^{8,9}.

DNA fluorescence-in-situ-hybridization (FISH) screening of male ES cells stably transfected with PEV3–GATA-1 showed localization of the transgene to the X chromosome in clone G4. The expected ratio in male to female progeny (38:34) was observed from 100% germline transmitting male chimaeras generated from this clone, F₁ females carrying the transgene and F₁ males (except one)

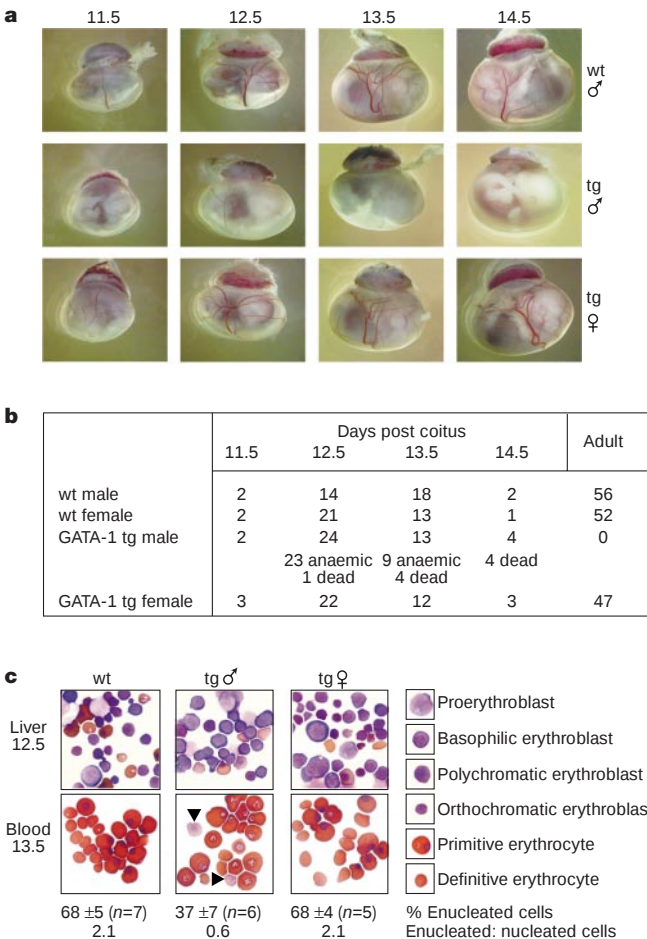


Figure 1 GATA-1 overexpression *in vivo*. **a**, Representative wild-type (wt) males and transgenics (tg) of either sex, 11.5 d.p.c. to 14.5 d.p.c., yolk sacs intact. **b**, Genotype and phenotype of embryos and adults, excluding F₁. Death was defined by lack of heartbeat, pallor was scored by anaemia. **c**, Disaggregated 12.5-d.p.c. fetal liver and 13.5-d.p.c. blood. Indicated by key is morphology of differentiating erythroid precursors and erythrocytes. Arrowheads indicate erythroid precursors in male transgenic blood (<3% of all cells). Percentage of enucleated erythrocytes (>300 events per embryo), number (*n*) of independent embryos and standard deviations, and ratio of enucleated to nucleated erythroid cells are shown below.

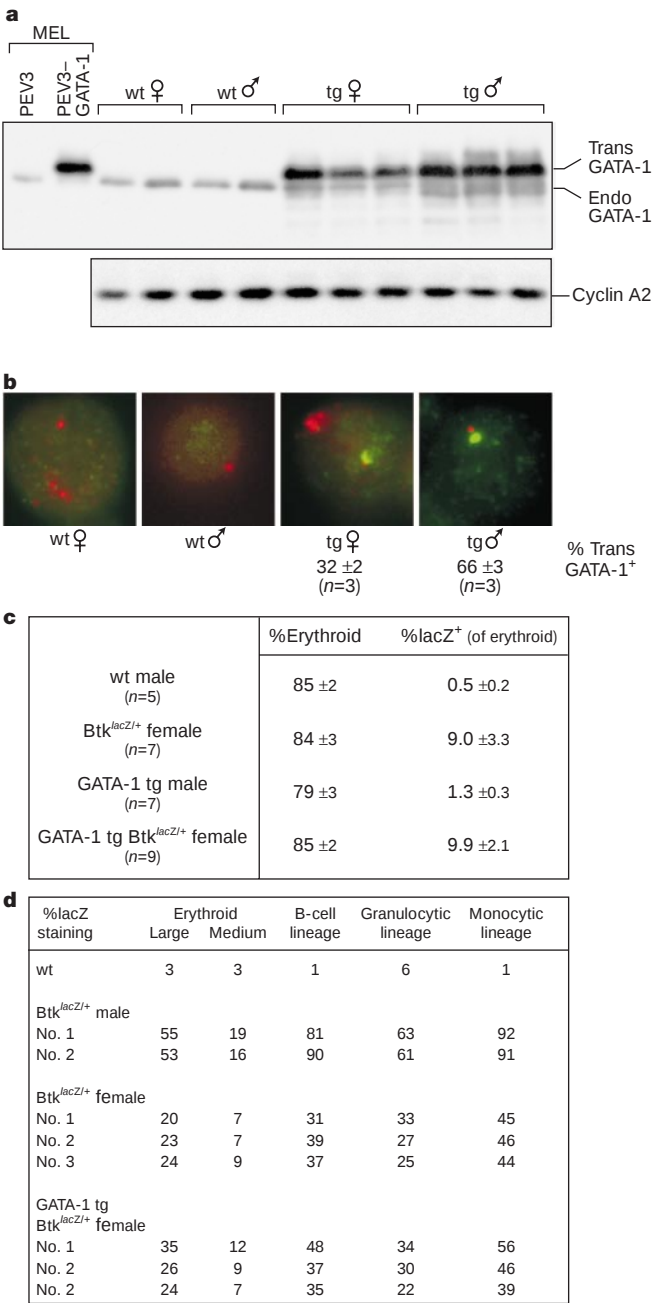


Figure 2 GATA-1 transgene is subject to X-inactivation. **a**, Western blot of nuclear extracts from 13.5-d.p.c. fetal livers, with anti-GATA-1 antibody N6. DMSO-induced control (PEV3) and GATA-1 overexpressing (PEV3–GATA-1) MEL cell samples are included. Positions of myc-tagged transgene-derived GATA-1 (trans GATA-1) and endogenous GATA-1 (endo GATA-1) are indicated. Cyclin A2 was used as a loading control. **b**, RNA FISH on 12.5-d.p.c. disaggregated fetal liver cells. Cells were stained for GATA-1 (green) and Xist (red). Representative cells are shown. The percentage of transgene-derived GATA-1 positive cells, number of embryos and standard deviations are indicated. **c**, Percentage of erythroid cells and percentage of lacZ⁺ erythroid cells in 13.5-d.p.c. fetal livers from female GATA-1 transgenic mice mated to Btk^{lacZ} males. Number of embryos and standard deviations are indicated. **d**, Percentage of lacZ⁺ staining cells in haematopoietic lineages of 6-week-old mice. Erythroid cells are arbitrarily divided into large and medium size according to the forward scatter (FSC) value. Note that GATA-1 transgenic Btk^{lacZ/+} female no. 1 which expresses lacZ in a high number of erythroid precursors also expresses lacZ in a higher than expected number of cells in the B-cell and monocytic lineages. This reflects the normal variance of the binomial distribution of X-inactivation balance in early pluripotent precursors.

being non-transgenic. The phenotypically male transgenic animal contained two X chromosomes (one carrying the transgene) and one Y chromosome (data not shown). Chimaeras generated from four other ES cell clones containing autosomal transgene integrations failed to give germline transmission.

We mated transgenic females and examined their progeny. Transgenic females were indistinguishable from wild-type littermates. Male transgenics were anaemic from 12.5 d.p.c. and dead by 14.5 d.p.c. (Fig. 1a, b). Back-crossing GATA-1 overexpressing females to FVB males for eight generations gave no viable transgenic males (excluding the XXY male). At 13.5 d.p.c., the number of enucleated erythrocytes relative to nucleated erythrocytes was reduced more than threefold in transgenic males compared with controls, showing that definitive erythropoiesis is inhibited (Fig. 1c). The fetal liver of transgenic male embryos, the site of definitive erythropoiesis at 13.5 d.p.c., was normal in size and cell number and contained similar numbers of apoptotic cells (terminal deoxynucleotidyl transferase (TdT)-mediated dUTP nick end labelling (TUNEL) assay; data not shown) compared with controls; however, transgenic male fetal livers contained significantly more early basophilic erythroid precursors and fewer late benzidine-positive cells (Fig. 1c), confirming that erythropoiesis is defective in fetal liver. Unexpectedly, female transgenics displayed no alteration in the morphology of cells in the fetal liver and had normal numbers of enucleated erythrocytes at this stage (Fig. 1c). Expression of putative GATA-1 target genes such as α -globin was unchanged in female transgenics, as RNA FISH detection of α -globin nascent transcripts on 13.5-d.p.c. fetal liver cells showed a normal number of cells transcribing α -globin (72% compared with 73% in controls). Furthermore, co-staining showed that half of these cells were positive for transgene-derived GATA-1 transcripts (data not shown), suggesting that the transgene is sensitive to X-inactivation.

Consistent with this interpretation, western blot analysis gave a higher level of transgene-derived GATA-1 protein in male versus female transgenic embryos (Fig. 2a). RNA FISH using a GATA-1 probe showed that at 12.5 d.p.c., 66% of transgenic male fetal liver cells were positive for transgene-derived GATA-1 nascent transcripts (appearing as a bright nuclear dot), whereas 32% of transgenic female cells were GATA-1 positive (Fig. 2b). This method did not detect endogenous GATA-1 transcripts. Co-staining for *Xist* RNA showed a single punctate signal in male fetal liver cells, corresponding to transcription from the *Xist* allele on the active X chromosome. Female cells also contained a larger area of accumulated *Xist* RNA on the inactive X chromosome (or Barr body)^{10,11}. Apparent co-localization of Barr body and GATA-1 signals occurred in less than 5% of double positive cells (Fig. 2b, and data not shown), confirming that male transgenics express the transgene pancellularly and females heterocellularly in response to X-inactivation. Survival of transgenic mice was not dependent on being female *per se*, as a transgenic XXY male (where X-inactivation occurs⁷) was

viable. One might argue that the observed phenotype arises from the X-chromosomal integration event itself; however, overexpression of the tamoxifen-inducible GATA-1-LBD in mice inhibited erythroid differentiation independent of the integration site. GATA-1-LBD accumulated to lower levels than the protein derived from PEV3-GATA-1 and consequently induced a milder phenotype (15% reduction compared with a 60% reduction in the proportion of differentiated erythroid cells in a CFU-E assay, data not shown).

To examine the relative contribution of GATA-1-overexpressing cells to the erythroid lineage, we bred transgenic females with males containing a *lacZ* gene insertion in the X-inactivation sensitive Bruton's tyrosine kinase (*Btk^{lacZ}*) locus¹². We found the same number of *lacZ*⁺ erythroid cells in the fetal livers of 13.5 d.p.c. *Btk^{lacZ}* and compound *Btk^{lacZ}/GATA-1* transgenic females (Fig. 2c). Adult *Btk^{lacZ}* males or females expressed *lacZ* in 50% or 25% of large erythroid precursors in the bone marrow, respectively. Compound *Btk^{lacZ}/GATA-1* transgenic females also expressed *lacZ* in 25% of large erythroid precursors (Fig. 2d). Thus, there is a normal representation of cells with an active *lacZ* gene in the erythroid compartment of GATA-1-overexpressing females throughout development. We conclude that there is no selection for or against GATA-1-overexpressing cells in females. As these females generate normal numbers of definitive erythrocytes, the GATA-1-overexpressing cells must be differentiating normally *in*

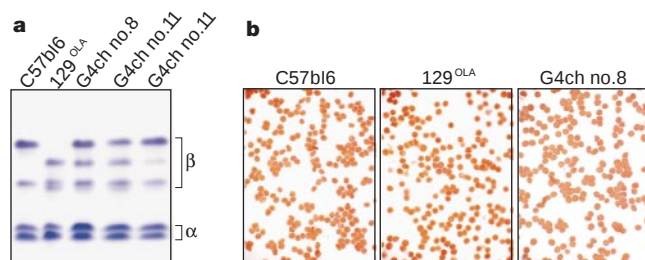


Figure 3 Cells overexpressing GATA-1 contribute to erythroid lineage in adult chimaeric mice. **a**, Triton-acid-urea gel of lysed blood from C57bl6 (host-blastocyst, Hb single), 129OLA (ES cell strain, Hb diffuse) or clone G4-derived chimaeras (G4ch nos 8, 11 and 14). Indicated are the positions of the α -type (α) and β -type globin (β) chains. **b**, Morphology of the blood derived from C57bl6, 129OLA and G4 chimaera no. 8 mice.

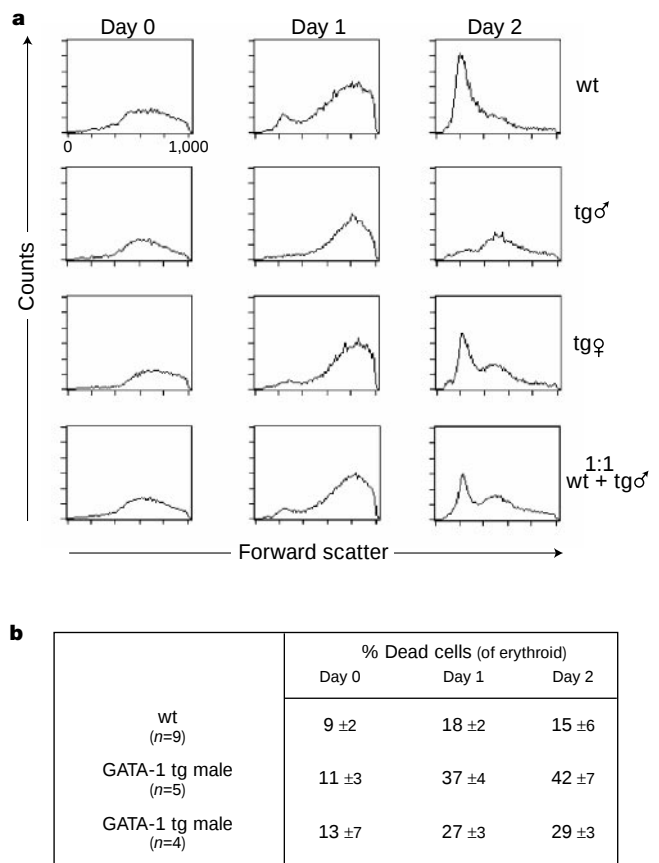


Figure 4 GATA-1 overexpressing CFU-Es fail to differentiate and undergo apoptosis *in vitro*. **a**, Representative histogram plots of FSC (indicating relative cell size) of viable erythroid cells from CFU-E assays performed on disaggregated fetal liver cells of 13.5-d.p.c. embryos. Cells harvested before (day 0), and after 24 h (day 1) and 48 h (day 2) of culture. Shown is the curve obtained by mixing transgenic male and wild-type samples in a 1:1 ratio after different periods of culture. **b**, Percentage of dead erythroid cells in same samples shown in **a**. Owing to a background of dead cells with a FSC of less than 327 (using our instrument settings), present in all CFU-E samples, cells with a FSC of less than 327 were excluded from this analysis.

vivo. To confirm that GATA-1-overexpressing cells contribute to the adult erythrocyte pool, chimaeric animals were generated with clone G4 cells (which are male and express the GATA-1 transgene pancellularly). These cells contributed up to 50% of the erythrocytes, as assayed by globin chain isoform analysis (Fig. 3a) and confirmed by glucose phosphate isomerase analysis (data not shown). These cells were morphologically normal (Fig. 3b). Furthermore, analysis of the blood of female transgenic mice showed no effect on other blood cell parameters, including red cell number, cell volume and haemoglobin content (data not shown).

To determine which differentiation stage is affected in the transgenic males, we carried out burst-forming unit-erythroid (BFU-E, reflecting early committed erythroid progenitors) and colony-forming unit-erythroid (CFU-E, reflecting a later stage in erythroid differentiation) assays. The same number of colonies in BFU-E and CFU-E assays were formed from both wild-type and transgenic 12.5-d.p.c. and 13.5-d.p.c. fetal livers (data not shown). CFU-E colonies develop from single cells¹³ and do not require accessory cells to form¹⁴, so the effect of GATA-1 overexpression on female transgenic precursors when no longer in close contact with non-overexpressing cells could be addressed. In male transgenic CFU-E colonies, the production of small late differentiated erythroid cells is inhibited (Fig. 4a), and after 24 h about 40% of erythroid cells were dead, compared with 20% in wild-type cultures (Fig. 4b). Female transgenic CFU-E cultures had an intermediate phenotype (Fig. 4a, b), identical to that found when wild-type and male transgenic cells were mixed before or after culture at a 1:1 ratio. This shows that female GATA-1-overexpressing erythroid precursors are intrinsically defective and behave identically to transgenic male erythroid precursors when removed from the fetal liver.

Gene function is 'cell autonomous' when a cell displays a phenotype that corresponds to its genotype, regardless of the genotype of surrounding cells. Gene function is defined as 'cell nonautonomous' when a cell exhibits a phenotype that does not correspond to its genotype¹⁵. In mosaics (that is, female transgenics and male chimaeras), the phenotype of GATA-1 transgenic cells is wild type. The function of the GATA-1 transgene is therefore cell

nonautonomous. The conventional interpretation would be that erythroid cells overexpressing GATA-1 are normal and that the GATA-1 transgene causes a defect in a non-erythroid cell normally supporting erythropoiesis. If so, then all erythroid precursors in female transgenics would behave identically *in vivo* and *in vitro*; however, this is not the case (Fig. 4). Thus, the conventional interpretation of cell nonautonomy is incorrect.

Perhaps the mutant erythroid cells produce a negative factor inhibiting differentiation, and this is diluted in mosaics. To be cell nonautonomous, however, expression of such a negative signalling factor would have to be the only defect in the cells overexpressing GATA-1. Considering the biochemical effects of GATA-1 overexpression in erythroid cells^{6,16,17}, this is unlikely. As GATA-1-overexpressing erythroid cells are intrinsically defective, they must be responding to a signal reversing these defects in mosaics. We therefore propose that wild-type cells produce a positive factor activating the differentiation of GATA-1 overexpressing cells. We tentatively name this activity 'red cell differentiation signal' or REDS. The cells producing REDS must be absent or reduced in transgenic males. Definitive erythropoiesis occurs in erythroblastic islands¹⁸. These contain a central macrophage surrounded by erythroid cells at all stages of maturation, with immature cells close to the macrophage and mature cells near the edge of the island^{18,19} (Fig. 5). The obvious candidate source of REDS is mature erythroid cells, as this is the only cell population clearly reduced in the transgenic males. Merely allowing wild-type fetal liver cells to contact cells overexpressing GATA-1 by mixing them in liquid culture does not restore differentiation to the latter (data not shown), indicating that disruption of the structure of the erythroblastic island also disrupts REDS activity. We suggest that mature erythroid cells on the periphery of the island are the source of REDS. In mosaics, late erythroid cells overexpressing GATA-1 are juxtaposed with REDS-producing wild-type mature cells in the same erythroblastic island and/or the neighbouring island. This overcomes the defects induced by high GATA-1 levels and allows the final stages of erythroid maturation to proceed. In male transgenic mice there is no such juxtaposition. Death ligands expressed by mature erythroid cells can induce the degradation of GATA-1

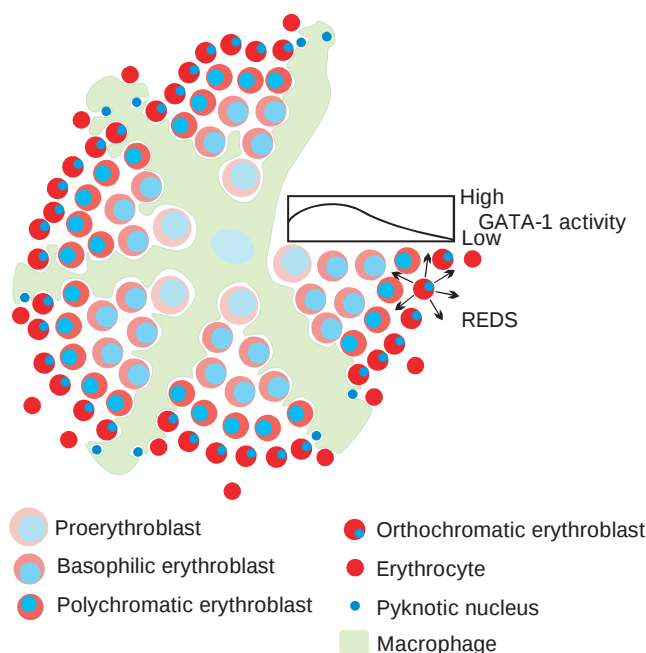


Figure 5 A simplified model of the erythroblastic island. The position of the central macrophage and the arrangement of the erythroid precursors (as indicated by the key) within the macrophage's cytoplasmic extensions are shown. The processes of pyknosis,

enucleation and phagocytosis of expelled pyknotic nuclei are represented. Arrows indicate the putative signal REDS, being produced by late differentiated erythroid cells. Graph indicates the proposed process of GATA-1 regulation late in the differentiation process.

(ref. 20). It is therefore possible that death ligands are a component of REDS. In support of this, death ligands partially reverse the effect of GATA-1 overexpression *in vitro* (data not shown).

There is one precedent for a signalling mechanism similar to the one that we propose. *Caenorhabditis elegans* larvae null for the IGF receptor homologue DAF-2 enter a state of diapause rather than developing into adulthood; however, DAF-2 null mosaic animals can become adult with all cells differentiating into adult tissues¹⁵. Significantly, an adult phenotype is not associated with an obligatory requirement for DAF-2 activity in a particular cell. Consequently, secondary signals may operate to ensure that all cells adopt the same developmental fate¹⁵. Furthermore, our observations have important implications in interpreting other cell-nonautonomous defects. For example, loss of the retinoblastoma protein results in a cell-nonautonomous inhibition of erythropoiesis. In accordance with the conventional interpretation of cell nonautonomy, it was concluded that retinoblastoma protein null erythroid cells are normal and that retinoblastoma protein function is required in stromal cells supporting erythropoiesis^{21,22}. This conclusion may be incorrect, as transplantation studies suggest that retinoblastoma protein is required in erythroid cells²³. This discrepancy remains unexplained; however, one resolution may be that the effect of retinoblastoma protein loss in erythroid cells is reversed by a signal supplied by wild-type cells, similar to the one that restores the normal differentiation of erythroid cells overexpressing GATA-1. □

Methods

Plasmids, probes and primers

Plasmids PEV3-GATA-1 and PEV3 have been described²⁴. Details of the cloning steps to produce the puromycin resistant version of PEV3-GATA-1 are available on request. Integration of PEV3-GATA-1 in ES cells and mice was screened by Southern blot using a 1-kilobase (kb) internal probe corresponding to the 5' end of the GATA-1 minigene. DNA FISH and RNA FISH was performed using a FITC-labelled 4.3-kb fragment spanning exon 2 to exon 6 of the GATA-1 gene, a Texas-Red labelled 21-kb human β -globin LCR probe and a Texas-Red labelled 6.5-kb fragment corresponding to the *Xist* RNA (a gift from N. Brockdorff). Embryos were sexed by PCR using primers specific for the *Zfy* gene²⁵.

Western blotting

Nuclear extracts were prepared as described²⁶. Each sample (5 μ g) was subject to electrophoresis through a 10% SDS polyacrylamide gel, transferred onto nitrocellulose and probed with anti-GATA-1 antibody N6 (Santa Cruz) or anti-cyclin A2 antibody C-19 (Santa Cruz) and an appropriate secondary antibody before detection using chemiluminescence.

ES cells

The puromycin resistant version of PEV3-GATA-1 was linearized with *Pvu*I and electroporated into 129^{OLA}-derived ES cells as described²⁷ and individual clones selected in 1 μ g ml⁻¹ puromycin.

Mice

Chimaeric mice were generated by injecting ES cell clones generated as above into C57Bl6 blastocysts. Chimeras were then bred with wild-type FVB males and screened by coat colour for germ-line transmission. Transgenic females were then mated to FVB males or Btk^{lacZ} males and sacrificed during gestation or allowed to go to term.

Blood and fetal liver cell analysis

Blood and/or disaggregated fetal livers were collected from embryos or adults and prepared on slides by cytocentrifugation. Slides were stained with neutral benzidine and a modified Giemsa-like stain.

Haemoglobin analysis

Globin chain representation in chimaeric animals was analysed as described²².

DNA and RNA FISH

DNA FISH was done as described²⁸. RNA FISH was performed on disaggregated fetal liver cells as described²⁹.

CFU-E assay

CFU-E assays were done as described³⁰. Fetal livers were disaggregated into single cells by passage through a 100- μ m mesh and plated at a density of 3×10^5 cells per ml in methylcellulose containing 1 unit ml⁻¹ Epo. Colonies were grown for times indicated and then collected and washed with PBS to remove residual methylcellulose before staining.

FACS analysis

Single-cell suspensions of bone marrow, fetal livers and cultured cells were prepared as above, stained and analysed by FACS as described³¹ with 5×10^4 events taken per sample. We used the following antibodies and stains: R-PE-conjugated TER119 antibody (Pharmingen), 7-aminoactinomycin-D (7AAD, Molecular Probes BV), fluorescein-di- β -D-galactopyranoside (FDG, Molecular Probes BV), biotin-conjugated ER-MP20 antibody, cyochrome-conjugated CD45R/B220 (Pharmingen) and Tricolor-streptavidin secondary antibody (Calatag Laboratories). Cell populations were divided as follows: non-viable (7AAD⁺), erythroid (TER119⁺), large erythroid (TER119⁺/FSC^{high}), medium-sized erythroid (TER119⁺/FSC^{medium}), B-cell lineage (B220⁺/ER-MP20⁺/TER⁻), granulocytic (ER-MP20⁺/TER119⁻) and monocytic (ER-MP20^{high}/TER119⁻).

Received 13 March; accepted 11 May 2000.

1. Pevny, L. *et al.* Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349**, 257–260 (1991).
2. Weiss, M. J., Keller, G. & Orkin, S. H. Novel insights into erythroid development revealed through *in vitro* differentiation of GATA-1⁻ embryonic stem cells. *Genes Dev.* **8**, 1184–1197 (1994).
3. Rossant, J. & Spence, A. Chimera and mosaics in mouse mutant analysis. *Trends Genet.* **14**, 358–363 (1998).
4. Weiss, M. J. & Orkin, S. H. Transcription factor GATA-1 permits survival and maturation of erythroid precursors by preventing apoptosis. *Proc. Natl Acad. Sci. USA* **92**, 9623–9627 (1995).
5. Takahashi, S. *et al.* Role of GATA-1 in proliferation and differentiation of definitive erythroid and megakaryocytic cells *in vivo*. *Blood* **92**, 434–442 (1998).
6. Whyatt, D. J. *et al.* The level of the tissue-specific factor GATA-1 affects the cell-cycle machinery. *Genes Funct.* **1**, 11–24 (1997).
7. Kuroda, M. I. & Meller, V. H. Transient *Xist*-ence. *Cell* **91**, 9–11 (1997).
8. Fujiwara, Y., Browne, C. P., Cunniff, K., Goff, S. C. & Orkin, S. H. Arrested development of embryonic red cell precursors in mouse embryos lacking transcription factor GATA-1. *Proc. Natl Acad. Sci. USA* **93**, 12355–12358 (1996).
9. Tsai, F. -Y., Browne, C. P. & Orkin, S. H. Knock-in mutation of transcription factor GATA-3 into the GATA-1 locus: partial rescue of GATA-1 loss of function in erythroid cells. *Dev. Biol.* **196**, 218–227 (1998).
10. Panning, B., Dausman, J. & Jaenisch, R. X chromosome inactivation is mediated by *Xist* RNA stabilization. *Cell* **90**, 907–916 (1997).
11. Sheardown, S. A. *et al.* Stabilization of *Xist* RNA mediates initiation of X chromosome inactivation. *Cell* **91**, 99–107 (1997).
12. Hendriks, R. W. *et al.* Inactivation of *Btk* by insertion by insertion of *lacZ* reveals defects in B cell development only past the pre-B cell stage. *EMBO J.* **15**, 4862–4872 (1996).
13. Cormack, D. Time-lapse characterization of erythrocytic colony-forming cells in plasma cultures. *Exp. Hematol.* **4**, 319–327 (1976).
14. Sawada, K., Krantz, S. B., Dessypris, E. N., Koury, S. T. & Sawyer, S. T. Human colony-forming units-erythroid do not require accessory cells but do require direct interaction with insulin-like growth factor 1 and/or insulin for erythroid development. *J. Clin. Invest.* **83**, 1701–1709 (1989).
15. Apfeld, J. & Kenyon, C. Cell nonautonomy of *C. elegans* *daf-2* function in the regulation of diapause and life span. *Cell* **95**, 199–210 (1998).
16. Dubart, A., Romeo, P. H., Vainchenker, W. & Dumenil, D. Constitutive expression of GATA-1 interferes with the cell-cycle regulation. *Blood* **87**, 3711–3721 (1996).
17. Briegel, K. *et al.* Regulation and function of transcription factor GATA-1 during red blood cell differentiation. *Development* **122**, 3839–3850 (1996).
18. Bessis, M., Lessin, L. S. & Beutler, E. In *Hematology* (eds Williams, W. J., Beutler, E., Erslev, A. J. & Lichtman, M. A.) 257–279 (McGraw-Hill, New York, 1983).
19. Bernard, J. The erythroblastic island: past and future. *Blood Cells* **17**, 5–14 (1991).
20. De Maria, R. *et al.* Negative regulation of erythropoiesis by caspase-mediated cleavage of GATA-1. *Nature* **401**, 489–493 (1999).
21. Williams, B. O. *et al.* Extensive contribution of Rb-deficient cells to adult chimeric mice with limited histopathological consequences. *EMBO J.* **13**, 4251–4259 (1994).
22. Maandag, E. C. *et al.* Developmental rescue of an embryonic-lethal mutation in the retinoblastoma gene in chimeric mice. *EMBO J.* **13**, 4260–4268 (1994).
23. Hu, N., Gulley, M. L., Kung, J. T. & Lee, E. Y.-H. Retinoblastoma gene deficiency has mitogenic but not tumorigenic effects on erythropoiesis. *Cancer Res.* **57**, 4123–4129 (1997).
24. Elefanty, A. G., Antoniou, M., Custodio, N., Carmo-Fonseca, M. & Grosfeld, F. G. GATA transcription factors associate with a novel class of nuclear bodies in erythroblasts and megakaryocytes. *EMBO J.* **15**, 319–333 (1996).
25. Nagamine, C. M., Chan, K. M., Kozak, C. A. & Lau, Y. F. Chromosome mapping and expression of a putative testis-determining gene in the mouse. *Science* **243**, 80–83 (1989).
26. Andrews, N. C. & Faller, D. V. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res.* **19**, 2499 (1991).
27. Pandolfi, P. P. *et al.* Targeted disruption of the GATA3 gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis. *Nature Genet.* **11**, 40–44 (1995).
28. Mulder, M. P. *et al.* Positional mapping of loci in the DiGeorge critical region at chromosome 22q11 using a new marker (D22S183). *Human Genet.* **96**, 133–141 (1995).
29. Wijgerde, M., Grosfeld, F. & Fraser, P. Transcription complex stability and chromatin dynamics *in vivo*. *Nature* **377**, 209–213 (1995).
30. Wong, P. M. C., Chung, S. W., Chui, D. H. K. & Eaves, C. J. Properties of the earliest clonogenic hemopoietic precursors to appear in the developing murine yolk sac. *Proc. Natl Acad. Sci. USA* **83**, 3851–3854 (1986).

Acknowledgements

We thank R. Bernards for his critical evaluation, discussions and materials contributing to this work. We also thank M. von Lindern, T. Verkerk, V. Lui, R. Delwel, S. Verbakel, G. Zafarana, N. Gillemans, D. Meijer, T. Stijnen, E. Dzierzak, E. Noteboom, R. Kerkhoven, J.-H. Dannenberg, H. te Riele and members of R. Bernards' laboratory for advice,

assistance and materials at various stages of the project. D.W. is supported by a long-term EMBO Fellowship. F.L. is supported by NWO (Netherlands). R.H. is supported by KNAW (Netherlands). NWO (Netherlands) and the EC supported this work.

Correspondence and requests for materials should be addressed to F.G. (e-mail: ggsveld@ch1.fgg.eur.nl).

Different initiation of pre-TCR and $\gamma\delta$ TCR signalling

Claude Saint-Ruf[†], Maddalena Panigada[‡], Orly Azogui[†], Pascale Debey[‡], Harald von Boehmer[§] & Fabio Grassi^{*†}

^{*} Institut Necker, U INSERM 373, Faculté de Médecine Necker, 156 Rue de Vaugirard, F75730, Paris cedex 15, France

[†] Unité INRA 806, Muséum National d'Histoire Naturelle, Institut de Biologie Physico-Chimique, 13, Rue Pierre et Marie Curie, 75005 Paris, France

[‡] Dipartimento di Biologia e Genetica per le Scienze Mediche, Università degli Studi di Milano, Via Viotti 3/5, 20133 Milano, Italy

[§] Harvard Medical School, Dana-Farber Cancer Institute, 44 Binney Street, Smith 736, Boston, Massachusetts, 02115, USA

Lineage choice is of great interest in developmental biology. In the immune system, the $\alpha\beta$ and $\gamma\delta$ lineages of T lymphocytes diverge during the course of the β -, γ - and δ -chain rearrangement of T-cell receptor (TCR) genes that takes place within the same precursor cell and which results in the formation of the $\gamma\delta$ TCR or pre-TCR proteins^{1–3}. The pre-TCR consists of the TCR β chain covalently linked to the pre-TCR α protein, which is present in immature but not in mature T cells which instead express the TCR α chain^{4,5}. Animals deficient in pre-TCR α have few $\alpha\beta$ lineage cells but an increased number of $\gamma\delta$ T cells. These $\gamma\delta$ T cells exhibit more extensive TCR β rearrangement than $\gamma\delta$ T cells from wild-type mice^{6,7}. These observations are consistent with the idea that different signals emanating from the $\gamma\delta$ TCR and pre-TCR instruct lineage commitment⁸. Here we show, by using confocal microscopy and biochemistry to analyse the initiation of signalling, that the pre-TCR but not the $\gamma\delta$ TCR colocalizes with the p56^{lck} Src kinase into glycolipid-enriched membrane domains (rafts) apparently without any need for ligation. This results in the phosphorylation of CD3 ϵ and Zap-70 signal transducing molecules. The results indicate clear differences between pre-TCR and $\gamma\delta$ TCR signalling.

The analysis by confocal microscopy of pre-TCR α (pT α) distribution in the plasma membrane of pre-TCR expressing thymocytes revealed a punctate staining pattern in contrast to the more diffuse and uniform distribution of CD25 (Fig. 1Aa–c). Peripheral $\alpha\beta$ T cells and $\gamma\delta$ thymocytes also exhibited a more uniform receptor staining (Fig. 1Ad–g). Quantitation of colocalization of pT α and TCR β chains on the surface of thymocytes revealed a strong positive correlation⁹ (correlation coefficient 0.95) and therefore the almost complete colocalization of the long pT α isoform with the TCR β chain in patches of the cell membrane from immature thymocytes (Fig. 1Ba–e). The patchy distribution of the pre-TCR was reminiscent of the distribution on the cell surface of the $\alpha\beta$ TCR after ligand binding¹⁰, which is caused by crosslinking of the $\alpha\beta$ TCR and subsequent localization into glycolipid-enriched membrane domains (rafts)^{11–14}. These rafts contain GPI-linked proteins, Src-like tyrosine kinases and other proteins targeted to rafts through saturated acyl chains^{15–18}. As the patchy distribution of the pre-TCR could conceivably be due to localization into rafts, we tested whether the pre-TCR, in the absence of any deliberate ligation, would colocalize with raft components in the membrane

of a SCID thymocyte-derived cell line (SCIET.27) that had been transfected with a productive TCR β gene and expressed the pre-TCR on the cell surface (SCB.29)⁴. Co-staining with GM1-binding cholera toxin B (CTxB) and pT α antibodies revealed the same clustered distribution of the pre-TCR as observed with thymocytes and colocalization with GM1 (Fig. 2Aa–d). A similar colocalization was also observed with thymocytes (data not shown). These results suggested that a putative pre-TCR ligand on thymic stroma was not essential for the patchy staining and colocalization with GM1.

The pre-TCR clustering was significantly reduced after cholesterol depletion with methyl- β -cyclodextrin (M β CD) (Fig. 2Ae, f), indicating the requirement for an intact lipid configuration. This suggests that the observed staining pattern was not caused by antibody crosslinking, because lateral mobility of transmembrane proteins is not inhibited by cholesterol depletion¹⁹. We also transfected the SCIET.27 cell line with a vector encoding a (V β 8.2D β 1J β 2.6C β 2) TCR β chain-enhanced green fluorescent (eGFP) fusion protein and found a clustered distribution of the tagged pre-TCR by confocal microscopy. No such clustering occurred with GFP protein alone (Fig. 2Ba–d). This further ruled out a requirement for pre-TCR crosslinking. (The large area of intracellular staining in Fig. 2Bc may represent internalized receptors).

To assess whether expression of a $\gamma\delta$ TCR would yield a similarly patchy staining, we transfected V γ 1J γ 1C γ 4 and V δ 7.1J δ 1C δ genes into the same pro-T cell line SCIET.27. Even though the $\gamma\delta$ TCR was expressed at similar levels, the staining was much more homogeneous (Fig. 2Be, f).

Genetic evidence implicates p56^{lck} as an essential component in pre-TCR function²⁰. We therefore tested whether the pre-TCR would colocalize with p56^{lck} in roughly the same way that it colocalizes with the $\alpha\beta$ TCR after crosslinking^{21–23}. We found that

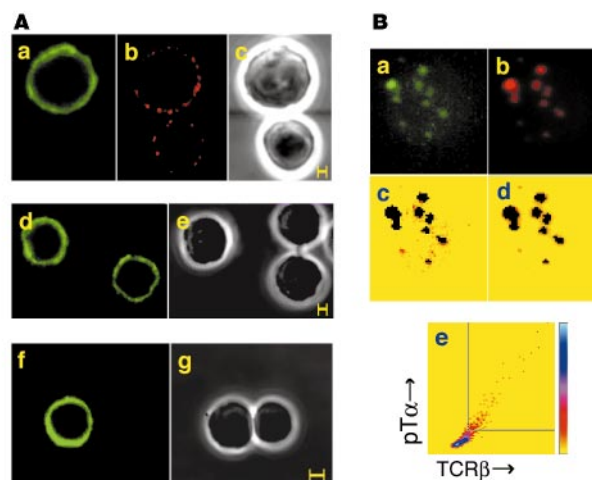


Figure 1 Analysis of the pre-TCR in the cell membrane of thymocytes. **A**, Punctate pattern of pT α chains on the surface of CD4⁺8[−] thymocytes. Co-staining of CD4⁺8[−] thymocytes with FITC-conjugated PC-61 (anti-CD25) monoclonal antibody (**a**) and biotinylated 2F5 (anti-pT α) monoclonal antibody revealed by streptavidin–TR (**b**), with phase contrast image (**c**). Staining of $\gamma\delta$ thymocytes with biotin-labelled GL-3 monoclonal antibody revealed by streptavidin–FITC (**d**, **e**). Staining of lymph nodes cells with FITC-conjugated H57-597 monoclonal antibody (**f**, **g**). Scale bar, 2 μ m. **B**, Quantitative colocalization analysis of TCR β and full-length pT α chains on the surface of CD4⁺8[−] thymocytes. Cells were stained with FITC-conjugated H57-597 (anti-TCR β) and biotin-labelled 2F5 (anti-pT α) monoclonal antibodies revealed by streptavidin–TR. Spatial distribution of TCR β (**a**) and pT α (**b**) stainings and corresponding SPIMAC 2D histogram (**e**). Areas outside the selected region in the 2D histogram represent the background; the pixels comprised in the region appear in black in the TCR β (**c**) and pT α (**d**) images and correspond to the green and red dots in **a** and **b**.

more than 80% of the pT α signals colocalized with p56^{lck} signals (correlation coefficient 0.85; Fig. 2Ca–e). The same analysis performed with pT α - and CD25-dependent fluorescence shows that pT α signals cannot be integrated into a homogeneous subset with respect to CD25 signals and that there is only occasional colocalization by chance (correlation coefficient 0.47; Fig. 2Cf–I).

The observed spontaneous colocalization of pre-TCR and p56^{lck} kinase in rafts might bring the pre-TCR-associated CD3 ϵ and CD3 ζ chains into contact with p56^{lck}, resulting in cell-autonomous activation of proximal signalling, that is, phosphorylation of CD3 ϵ and CD3 ζ chains as well as Zap-70. We tested this by directly analysing the colocalization biochemically, and by determining CD3 and Zap-70 phosphorylation in the pre-TCR competent (SCB.29) and pre-TCR incompetent (SCIET.27) cell lines. Immunoblot analysis with

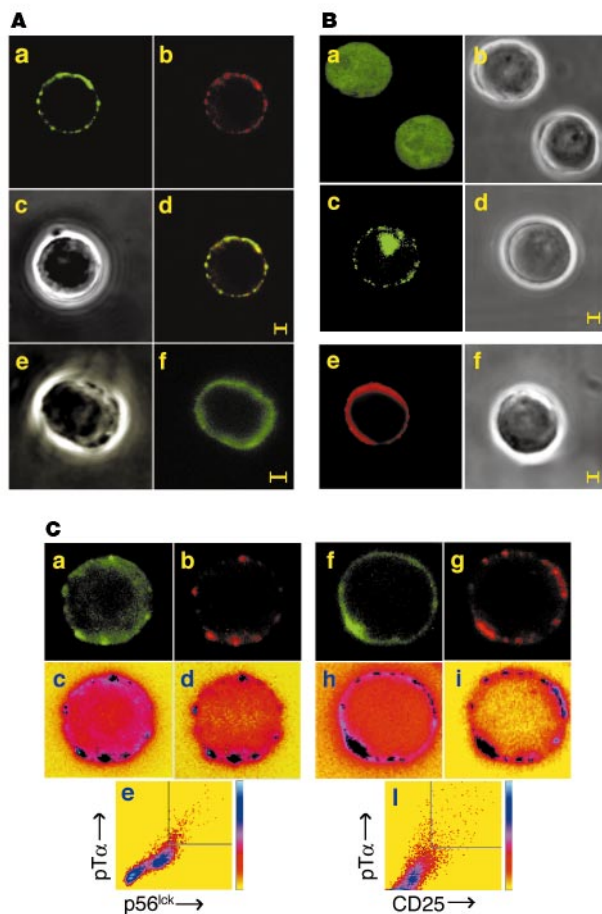


Figure 2 Analysis of the pre-TCR in the cell membrane of SCB.29 cells. **A**, Colocalization of GM₁ with pT α chains on the surface of SCB.29 cells by confocal microscopy. Cells were stained with FITC-labelled CTxB monoclonal antibody (a) and biotinylated 2F5 monoclonal antibody revealed by streptavidin–TR (b); phase contrast (c) and merged (d) images are shown. By quantitative colocalization analysis, 77% of pT α staining colocalizes with CTxB. Staining of SCB.29 cells treated with M β CD with biotinylated 2F5 monoclonal antibody revealed by streptavidin–FITC (e, f). Scale bar, 2 μ m. **B**, Analysis by confocal microscopy of SCIET.27 cells expressing eGFP (a, b) and TCR β –eGFP (c, d). Staining of SCIET.27 cells expressing $\gamma\delta$ TCR with biotin-labelled 3A10 (anti-TCR δ) monoclonal antibody revealed by streptavidin–TR (e, f). Scale bar, 2 μ m. **C**, Quantitative colocalization analysis of pT α chains and p56^{lck} in SCB.29 cells. Cells were stained with anti-Lck monoclonal antibody revealed by FITC-labelled goat anti-mouse Ig (a) and rabbit anti-pT α Ig revealed by TR-conjugated goat anti-rabbit Ig (b); the corresponding SPIMAC 2D histogram is shown (e). Areas outside the selected region in the histogram represent the background. The selected region is depicted in black in the p56^{lck} (c) and pT α (d) images. As a control, cells were stained with FITC-conjugated PC-61 (anti-CD25) monoclonal antibody (f) and rabbit anti-pT α Ig revealed by TR-conjugated goat anti-rabbit Ig (g).

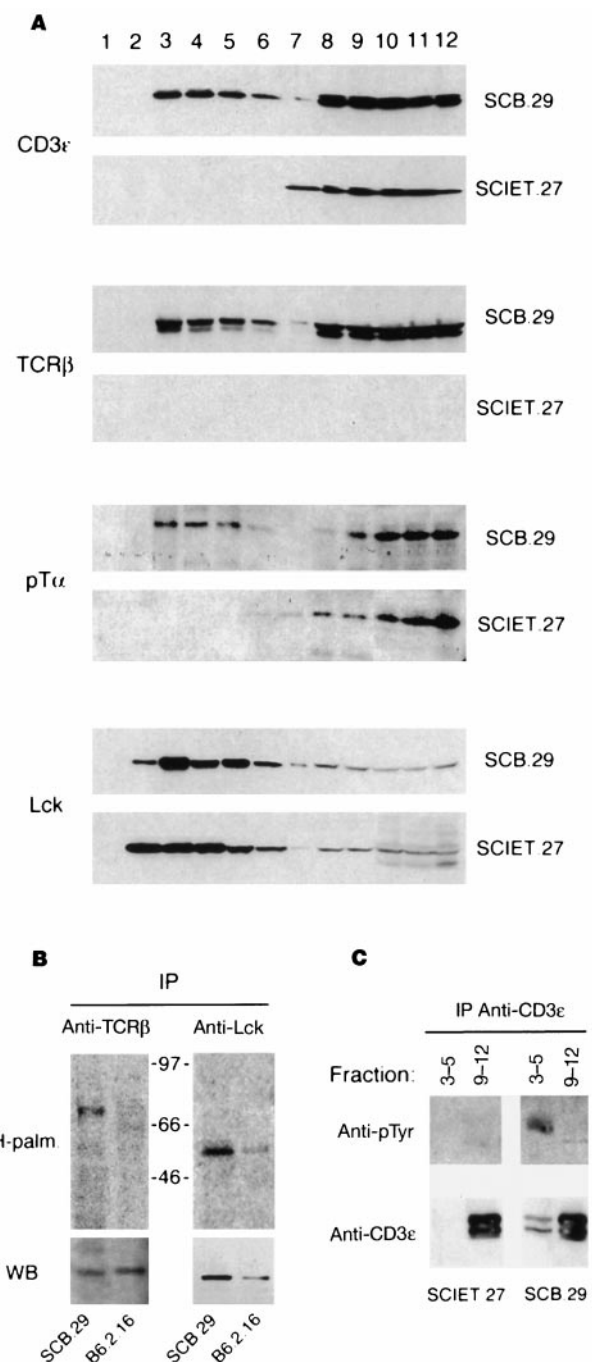


Figure 3 Biochemical analysis of the pre-TCR. **A**, Partition of pre-TCR into rafts analysed by sucrose gradient fractionation. Western blot with the indicated antibodies of TCA-precipitated sucrose gradient fractions from SCB.29 and SCIET.27 cell lysates. The analysis with the different antibodies was performed on the same membranes for both SCB.29 and SCIET.27 cell lysates. **B**, Palmitoylation of pT α chain. Lysates from biosynthetically labelled cells with ³H-palmitate were immunoprecipitated either with anti-TCR β or anti-Lck monoclonal antibodies. After separation in non-reducing gels, the immunoprecipitates were transferred onto nitrocellulose membranes. Autoradiographies (4-week exposure) and corresponding western blot with F23.1 and anti-Lck monoclonal antibodies are shown. **C**, CD3 ϵ phosphorylation in raft fractions of SCB.29 cells. Anti-CD3 ϵ immunoprecipitates of the indicated fractions from sucrose gradients were resolved in non-reducing gels, transferred to nitrocellulose membranes and sequentially probed with anti-phosphotyrosine (4G10) and anti-CD3 ϵ antibodies. In non-reducing conditions, CD3 ϵ chains migrate as a doublet and tyrosine phosphorylation is confined to the upper band.

specific antibodies of sucrose gradient fractions containing Triton X-100 SCB.29 and SCIET.27 cell lysates revealed the presence of pT α and CD3 ϵ chains in low-density fractions in SCB.29 but not SCIET.27 cells (Fig. 3A). In contrast, p56^{lck} was found in low-density fractions of both cells owing to its spontaneous localization into rafts^{16,17}. The recovery of TCR β , pT α and CD3 ϵ proteins from low-density fractions was abolished by treatment with M β CD before the lysis of cells (data not shown).

These biochemical results show independently that the pre-TCR localizes cell-autonomously into rafts. Such localization could possibly be mediated by the palmitoylation of the juxtamembraneous cysteine residue conserved in both the human and the murine pT α protein²⁴. We therefore used metabolic labelling with ³H-palmitate, followed by immunoprecipitation with TCR β antibodies from cell lysates of SCB.29 and the B6.2.16 hybridoma (which expressed the same TCR β chain as SCB.29, but in covalent linkage with the TCR α chain). Whereas we did not detect labelled $\alpha\beta$ TCR, we observed incorporation of ³H-palmitate into the pre-TCR in the SCB.29 cells (Fig. 3B). The ability to abolish the label with hydroxylamine confirmed the linkage of the ³H-palmitate label to the protein. The palmitoylated band corresponding to the pre-TCR was detected also when immunoprecipitations were carried out with anti-pT α and anti-CD3 ϵ antibodies (data not shown).

The recovery of phosphorylated CD3 ϵ chains from low-density fractions of cell lysates from SCB.29 but not SCIET.27 cells (Fig. 3C) suggested a functional significance of pre-TCR and p56^{lck} colocalization in rafts. Previously, weakly phosphorylated ζ -chains were found to coprecipitate with Zap-70 in unstimulated SCB.29 cells,

suggesting activation of Zap-70 (ref. 25). Zap-70 is recruited and activated by phosphorylated CD3 ϵ (ref. 26). We therefore analysed whether phosphorylated Zap-70 could be recovered from both particulate and soluble fractions of SCB.29 cells but not SCIET.27 cells (Fig. 4A). Phosphorylation was detected in SCIET.27 cells that were independently transfected with five different TCR β genes (V β 8.2 D β 1J β 2.6C β 2, V β 8.2 D β 2 J β 2.6C β 2, two V β 12 J β 2.6C β 2 with different CDR3 regions, and V β 4J β 2.6C β 2) (data not shown), but not in non-transfected cells. This effectively eliminates the possibility that phosphorylation was due to the selection of a rare variant in culture and/or being dependent on a particular TCR β /pT α combination. Disruption of raft integrity by M β CD or inhibition of p56^{lck} by the specific inhibitor PP2 resulted in complete loss of Zap-70 phosphorylation (Fig. 4A). The phosphorylation in SCB.29 cells was unlikely to be caused merely by elevated levels of CD3 and Zap-70 after expression of the pre-TCR: whereas particular $\gamma\delta$ TCR and TCR β transfectants of SCIET.27 exhibited the same CD3 ϵ expression on the cell surface by cytometry, phosphorylated Zap-70 was detected in TCR β but not TCR $\gamma\delta$ transfectants (Fig. 4B), even though Zap-70 was detected at comparable levels in the $\gamma\delta$ TCR-expressing cell lines by immunoblotting and was targeted to the plasma membrane independently of its phosphorylation, as previously described²⁷ (Fig. 4B).

Thus, a fundamental difference exists in the manner by which the $\gamma\delta$ TCR and the pre-TCR initiate signal transduction. Our data indicate that the pre-TCR α protein may have specifically evolved to pair with the TCR β chain and to segregate cell-autonomously into rafts as do other palmitoylated proteins. This sequestration into rafts apparently suffices to initiate the pre-TCR signalling. Such signalling may or may not involve binding of the pre-TCR to a ligand expressed by pre-T cells themselves. However, data obtained with a truncated pre-TCR may suggest that binding to the extracellular portion of the pre-TCR is not required²⁸. Differences in pre-TCR and $\gamma\delta$ TCR signalling may instruct immature T cells to develop along different pathways. □

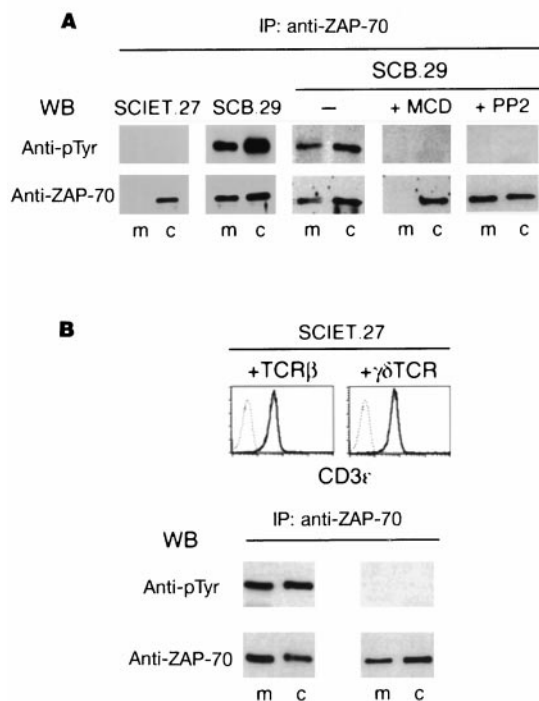


Figure 4 Cell-autonomous pre-TCR signalling in rafts. **A**, Constitutive phosphorylation of Zap-70 in SCB.29 cells, and dependence on raft integrity and p56^{lck} activity. Membrane (m) and cytosol (c) fractions from SCIET.27 and SCB.29 cells were immunoprecipitated with anti-ZAP-70 antibodies and analysed with anti-phosphotyrosine (4G10) monoclonal antibody and anti-ZAP-70 rabbit Ig; the same analysis on untreated and M β CD- and PP2-treated SCB.29 cells is shown. **B**, Lack of constitutive Zap-70 phosphorylation by $\gamma\delta$ TCR expression in SCIET.27 cells. Flow cytometry of two clones from TCR β and $\gamma\delta$ TCR transfections of SCIET.27 cells stained with biotin-labelled 145-2C11 (anti-CD3 ϵ) monoclonal antibody revealed by streptavidin-PE. The same two clones were analysed for Zap-70 phosphorylation in both membrane (m) and cytosol (c) fractions.

Methods

Antibodies and expression vectors

We used the following antibodies: FITC-conjugated anti-TCR β (H57-597) and anti-CD25 (PC-61) monoclonal antibodies (Pharmingen); biotin-conjugated anti-pT α (2F5)⁷, anti-TCR δ (3A10), anti-CD3 ϵ (145.2C11) and anti-TCR $\gamma\delta$ (GL-3) (Pharmingen) monoclonal antibodies; unlabelled anti-CD4 (RL1.72), anti-CD8 (31LM), anti-CD3((145.2C11) and anti-V β 8 (F23.1) monoclonal antibodies purified from culture supernatants; anti-Lck (MOL171) (Pharmingen) and anti-phosphotyrosine (4G10) monoclonal antibody (Upstate Biotechnology); rabbit anti-pT α ⁷, anti-Lck, anti-Zap-70 and goat anti-CD3 ϵ immunoglobulins (Ig) (Santa Cruz). We used the following secondary reagents: streptavidin-labelled with FITC, phycoerythrin (PE), Texas red (TR) (Southern Biotechnology) and horseradish peroxidase (HRP) (Amersham), TR-conjugated goat anti-rabbit and FITC-conjugated goat anti-mouse Ig (Southern Biotechnology), sheep anti-mouse, donkey anti-rabbit (Amersham) and rabbit anti-goat (Valbiotech) Ig labelled with HRP. For the generation of SCIET.27 cells expressing the pre-TCR tagged with enhanced green fluorescent protein (eGFP), full-length cDNA of the TCR β gene V β 8.2D β 1J β 2.6C β 2 was cloned in pEGFP-N1 vector (Clontech) to express TCR β fused to the amino terminus of eGFP. For transfection of SCIET.27 cells with TCR β , rearranged TCR β DNAs were cloned into pHb Apr-1-neo vector²⁹, whereas for expression of TCR $\gamma\delta$, SCIET.27 cells were co-transfected with TCR γ and TCR δ genes derived from cDNA of TCR $\gamma\delta$ transgenic thymocytes and cloned into pHb Apr-1-neo and pSVL (Pharmacia) plasmids, respectively. SCIET.27 cells were transfected by electroporation, and surface expressions of the pre-TCR or TCR $\gamma\delta$ on resistant clones were checked in flow cytometry with appropriate monoclonal antibodies.

Fluorescence microscopy and image analysis

Double negative CD4⁺8⁺ thymocytes expressing the pre-TCR were highly enriched from thymi of 3–4-week-old C57BL/6 TCR α ^{-/-} mice by complement-mediated depletion of CD4 and CD8 expressing thymocytes followed by selection at the cell sorter of cells negative for DX 5 (NK cells) and $\gamma\delta$ TCR expression. For staining of $\gamma\delta$ thymocytes, the same procedure was followed with exclusion of the anti- $\gamma\delta$ TCR monoclonal antibody during sample preparation for cell sorting. Peripheral lymph node T cells were enriched from lymph nodes by Dynabead (Dyna) depletion of IgM⁺ B cells. The SCIET.27 and SCB.29 cell lines have been described⁴. After adhesion on poly-L-lysine coated coverslip, the cells were fixed in paraformaldehyde and incubated on ice for 45 min with antibodies in PBS, 3% fetal calf serum (FCS) and sodium azide. For staining of the gangliosides GM₁,

FITC-labelled cholera toxin B (CTxB) subunit (Sigma) was used in the absence of FCS. When required, washed cells were stained with the specific second step reagent.

For intracellular staining, fixed cells were labelled with anti-pT α antibodies and permeabilized in PBS with 0.15% Triton X-100 for 5 min followed by staining with anti-p56^{lck} antibodies. Cells were depleted of membrane cholesterol by treatment with 10 mM methyl- β -cyclodextrin (Sigma) in Dulbecco's modified Eagle's medium supplemented with 0.4% fatty-acid-free bovine serum albumin, 1 μ g ml⁻¹ transferrin, 8.1 μ g ml⁻¹ monothio glycerol and 20 mM HEPES pH 7.4 (lipid-free medium) for 1 h at 37 °C. Confocal microscopy was done on a LSM 510 Zeiss confocal microscope. For quantitative colocalization analysis, we used a Zeiss inverted microscope (Axiovert 35) as conventional microscopy is more sensitive than confocal microscopy for this analysis⁹. Images were acquired and processed with a CCD camera (Photometrics) cooled to 10 °C and coupled to the IPLAB Spectrum Imaging software. The spatial distribution of two labelled molecules was analysed on recorded images using the Co-localization Image Analysis program from the SPIMAC (Spectral Imaging on Macintosh) software as described⁹.

Subcellular fractionations, biosynthetic labelling and western blotting

Cells were lysed in Triton X-100 as described¹², mixed with 1 ml 80% sucrose, and overlaid with two phases of 2 ml 30% sucrose and 1 ml 5% sucrose, respectively. Samples were centrifuged at 200,000g at 4 °C for 14–16 h, and 0.4 ml fractions were collected and numbered from the top of the gradient, precipitated in trichloroacetic acid (TCA), resolved by SDS–PAGE in reducing 8–15% gradient gels, transferred to nitrocellulose membranes and blotted with the indicated antibodies and specific secondary reagents. For analysis of CD3 ϵ phosphorylation, raft fractions (from 3 to 5) and loading zone fractions (from 9 to 12) were immunoprecipitated with anti-CD3 ϵ monoclonal antibody followed by Protein A Sepharose. Anti-CD3 ϵ immunoprecipitates were resolved by SDS–PAGE in non-reducing 8–15% gradient gels, transferred to nitrocellulose membranes and sequentially blotted with anti-phosphotyrosine mAb 4G10 and goat anti-CD3 ϵ Ig.

Biosynthetic labelling of endogenous palmitate was carried out by incubation of cells in the presence of ³H-palmitate (0.5 mCi) for 4 h at 37 °C in lipid-free medium without monothio glycerol. Cells were lysed in 1% Triton X-100, 0.2% saponin, and lysates were immunoprecipitated with either anti-TCR β or anti-Lck monoclonal antibody coupled to Protein A Sepharose. After separation by SDS–PAGE in non-reducing conditions, samples were transferred to nitrocellulose membranes, which were used for autoradiography with BioMax TranScreen intensifying system (Kodak). For membrane and cytosol separation, either untreated cells or cells treated with 10 mM M β CD or 10 μ M PP2 (Calbiochem) for 15 min at 37 °C, were lysed in hypotonic buffer (20 mM Tris-HCl pH 7.5, 1 mM EGTA, 1 mM MgCl₂, 0.5 mM DTT and protease inhibitors) and disrupted by homogenization on ice with a Dounce homogenizer. Salt concentration was adjusted to 150 mM NaCl, and intact cells, nuclei and cytoskeleton were pelleted by two centrifugations at 5,000 r.p.m. for 5 min in microcentrifuge at 4 °C. The supernatant was centrifuged at 100,000g for 1 h and membrane pellets solubilized in 0.5% Triton X-100 for immunoprecipitation with anti-ZAP-70 Ig. The supernatant corresponding to cytosol was equilibrated to 0.5% Triton X-100 and immunoprecipitated with anti-ZAP-70 Ig.

Received 30 March; accepted 25 May 2000.

1. von Boehmer, H. & Fehling, H. J. Structure and function of the pre-T cell receptor. *Annu. Rev. Immunol.* **15**, 433–452 (1997).
2. Godfrey, D. I., Kennedy, J., Suda, T. & Zlotnik, A. A developmental pathway involving four phenotypically and functionally distinct subsets of CD3⁺CD4⁺CD8⁺ triple negative adult mouse thymocytes defined by CD44 and CD25 expression. *J. Immunol.* **150**, 4244–4252 (1993).
3. Rodewald, H. R. *et al.* Fc γ RIII/III and CD2 expression mark distinct subpopulations of immature CD4⁺ murine thymocytes: In vivo developmental kinetics and T cell receptor β chain rearrangement status. *J. Exp. Med.* **177**, 1079–1092 (1992).
4. Groettrup, M. *et al.* A novel disulfide-linked heterodimer on pre-T cells consists of the T cell receptor β chain and a 33 kD glycoprotein. *Cell* **75**, 283–294 (1993).
5. Saint-Ruf, C. *et al.* Analysis and expression of a cloned pre-T cell receptor gene. *Science* **266**, 1208–1212 (1994).
6. Fehling, H. J., Krotkova, A., Saint-Ruf, C. & von Boehmer, H. Crucial role of the pre-T-cell receptor α gene in development of $\alpha\beta$ but not $\gamma\delta$ T cells. *Nature* **375**, 795–798 (1995).
7. Aifantis, I. *et al.* On the role of the pre-T cell receptor in $\alpha\beta$ versus $\gamma\delta$ lineage commitment. *Immunity* **9**, 649–655 (1998).
8. von Boehmer, H. *et al.* Crucial function of the pre-T cell receptor (TCR) in TCR β selection, TCR β allelic exclusion and $\alpha\beta$ versus $\gamma\delta$ lineage commitment. *Immunol. Rev.* **165**, 111–119 (1998).
9. Amirand, C. *et al.* Three distinct sub-nuclear populations of HMG-I protein of different properties revealed by co-localization image analysis. *J. Cell Sci.* **111**, 3551–3561 (1998).
10. James, P. W., Ley, S. C. & Magee, A. I. Aggregation of lipid rafts accompanies signaling via the T cell antigen receptor. *J. Cell Biol.* **147**, 447–461 (1999).
11. Xavier, R., Brennan, T., Li, Q., McCormack, C. & Seed, B. Membrane compartmentation is required for efficient T cell activation. *Immunity* **8**, 723–732 (1998).
12. Montixi, C. *et al.* Engagement of T cell receptor triggers its recruitment to low-density detergent-insoluble membrane domains. *EMBO J.* **17**, 5334–5348 (1998).
13. Simons, K. & Ikonen, E. Functional rafts in cell membranes. *Nature* **387**, 569–572 (1997).
14. Harder, T., Scheiffele, P., Verkade, P. & Simons, K. Lipid domain structure of the plasma membrane revealed by patching of membrane components. *J. Cell Biol.* **141**, 929–942 (1998).
15. Brown, D. A. & Rose, J. K. Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface. *Cell* **68**, 533–544 (1992).
16. Shenoy-Scaria, A. M., Dietzen, D. J., Kwong, J., Link, D. C. & Lublin, D. M. Cysteine 3 of Src family protein tyrosine kinase determines palmitoylation and localization in caveolae. *J. Cell Biol.* **126**, 353–363 (1994).
17. Rodgers, W., Crise, B. & Rose, J. K. Signals determining protein tyrosine kinase and glycosyl-phosphatidylinositol-anchored protein targeting to a glycolipid-enriched membrane fraction. *Mol. Cell Biol.* **14**, 5384–5391 (1994).

18. Zhang, W., Triple, R. P. & Samelson, L. E. LAT palmitoylation: its essential role in membrane microdomain targeting and tyrosine phosphorylation during T cell activation. *Immunity* **9**, 239–246 (1998).
19. Sheets, E. D., Holowka, D. & Baird, B. Critical role for cholesterol in Lyn-mediated tyrosine phosphorylation of FceRI and their association with detergent-resistant membranes. *J. Cell Biol.* **145**, 877–887 (1999).
20. Anderson, S. J. & Perlmutter, R. M. A signaling pathway governing early thymocyte maturation. *Immunol. Today* **16**, 99–105 (1995).
21. Monks, C. R., Freiberg, B. A., Kupfer, H., Sciaky, N. & Kupfer, A. Three-dimensional segregation of supramolecular activation clusters in T cells. *Nature* **395**, 82–86 (1998).
22. Brown, D. A. & London, E. Functions of lipid rafts in biological membranes. *Annu. Rev. Cell Dev. Biol.* **14**, 111–136 (1998).
23. Lanzavecchia, A., Lezzi, G. & Viola, A. From TCR engagement to T cell activation: a kinetic view of T cell behavior. *Cell* **96**, 1–4 (1999).
24. Del Porto, P., Bruno, L., Mattei, M. G., von Boehmer, H. & Saint-Ruf, C. Cloning and comparative analysis of the human pre-T-cell receptor α -chain gene. *Proc. Natl Acad. Sci. USA* **92**, 12105–12109 (1995).
25. van Oers, N. S., von Boehmer, H. & Weiss, A. The pre-T cell receptor (TCR) complex is functionally coupled to the TCR- ζ subunit. *J. Exp. Med.* **182**, 1585–1590 (1995).
26. Wange, R. L., Malek, S. N., Desiderio, S. & Samelson, L. E. Tandem SH2 domains of ZAP-70 bind to T cell antigen receptor ζ and CD3 ϵ from activated Jurkat T cells. *J. Biol. Chem.* **268**, 19797–19801 (1993).
27. Huby, R. D. J., Iwashima, M., Weiss, A. & Ley, S. C. ZAP-70 protein tyrosine kinase is constitutively targeted to the T cell cortex independently of its SH2 domains. *J. Cell Biol.* **137**, 1639–1649 (1997).
28. Irving, B. A., Alt, F. W. & Killeen, N. Thymocyte development in the absence of pre-T cell receptor extracellular immunoglobulin domains. *Science* **280**, 905–908 (1998).
29. Gunning, P., Leavitt, J., Muscat, G., Ng, S. Y. & Keddes, L. A human β -actin expression vector system directs high-level accumulation of antisense transcripts. *Proc. Natl Acad. Sci. USA* **84**, 4831–4835 (1987).

Acknowledgements

We thank C. Amirand for advice in quantitative colocalization analysis; J. Feinberg, C. Garcia and I. Aifantis for help in thymocyte analysis; E. Barbier for advice on biochemistry; P. Pereira for $\gamma\delta$ TCR transgenic mice; and Y. Goureau for assistance in confocal microscopy. We thank E. D. Smith for help with the artwork and L. Holcomb for preparation of the manuscript. H.v.B. is supported by the Koerber Foundation (Germany). E.G. is a recipient of a Biotech grant from European Community.

Correspondence and requests for materials should be addressed to H.v.B. (e-mail Harald_von_Boehmer@dfci.harvard.edu).

Distinct β -catenins mediate adhesion and signalling functions in *C. elegans*

Hendrik C. Korswagen*, Michael A. Herman† & Hans C. Clevers*‡

* Department of Immunology and ‡ Center for Biomedical Genetics, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands
† Program in Molecular, Cellular and Developmental Biology, Division of Biology, Kansas State University, Manhattan, Kansas 66506, USA

In flies and vertebrates, Armadillo/ β -catenin forms a complex with Tcf/Lef-1 transcription factors, serving as an essential co-activator to mediate Wnt signalling. It also associates with cadherins to mediate adhesion. In *Caenorhabditis elegans*, three putative β -catenin homologues have been identified: WRM-1, BAR-1 and HMP-2. WRM-1 and the Tcf homologue POP-1 mediate Wnt signalling by a mechanism that has challenged current views of the Wnt pathway^{1–3}. Here we show that BAR-1 is the only β -catenin homologue that interacts directly with POP-1. BAR-1 mediates Wnt signalling by forming a BAR-1/POP-1 bipartite transcription factor that activates expression of Wnt target genes such as the Hox gene *mab-5*. HMP-2 is the only β -catenin homologue that interacts with the single cadherin of *C. elegans*, HMR-1. We conclude that a canonical Wnt pathway exists in *C. elegans*. Furthermore, our analysis shows that the functions of *C. elegans* β -catenins in adhesion and in signalling are performed by separate proteins.

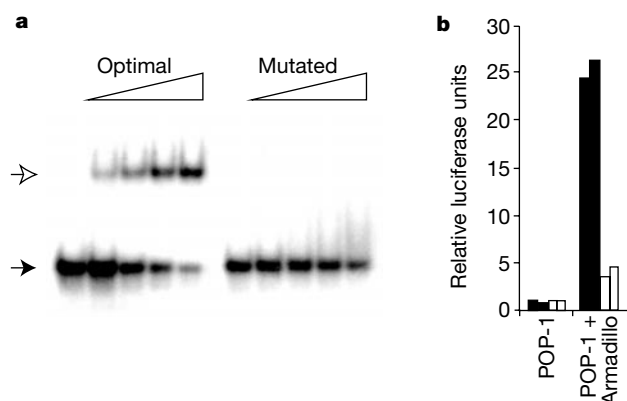


Figure 1 POP-1 is a functional Tcf homologue. **a**, Gel retardation analysis was performed with recombinant POP-1(180–280) containing the HMG DNA-binding domain. The POP-1 HMG domain binds the optimal Tcf target sequence GATCAAAG, but not the mutated sequence GATCCCCG. White arrow, shifted band; black arrow, free probe. **b**, POP-1 and

Armadillo activate a Tcf reporter gene. Expression constructs for POP-1 and Armadillo were transfected into IIAI.6 cells with the TOP luciferase reporter¹⁸ which contains optimal Tcf-binding sites (black bars), or the negative control FOP luciferase which contains mutated Tcf-binding sites (white bars).

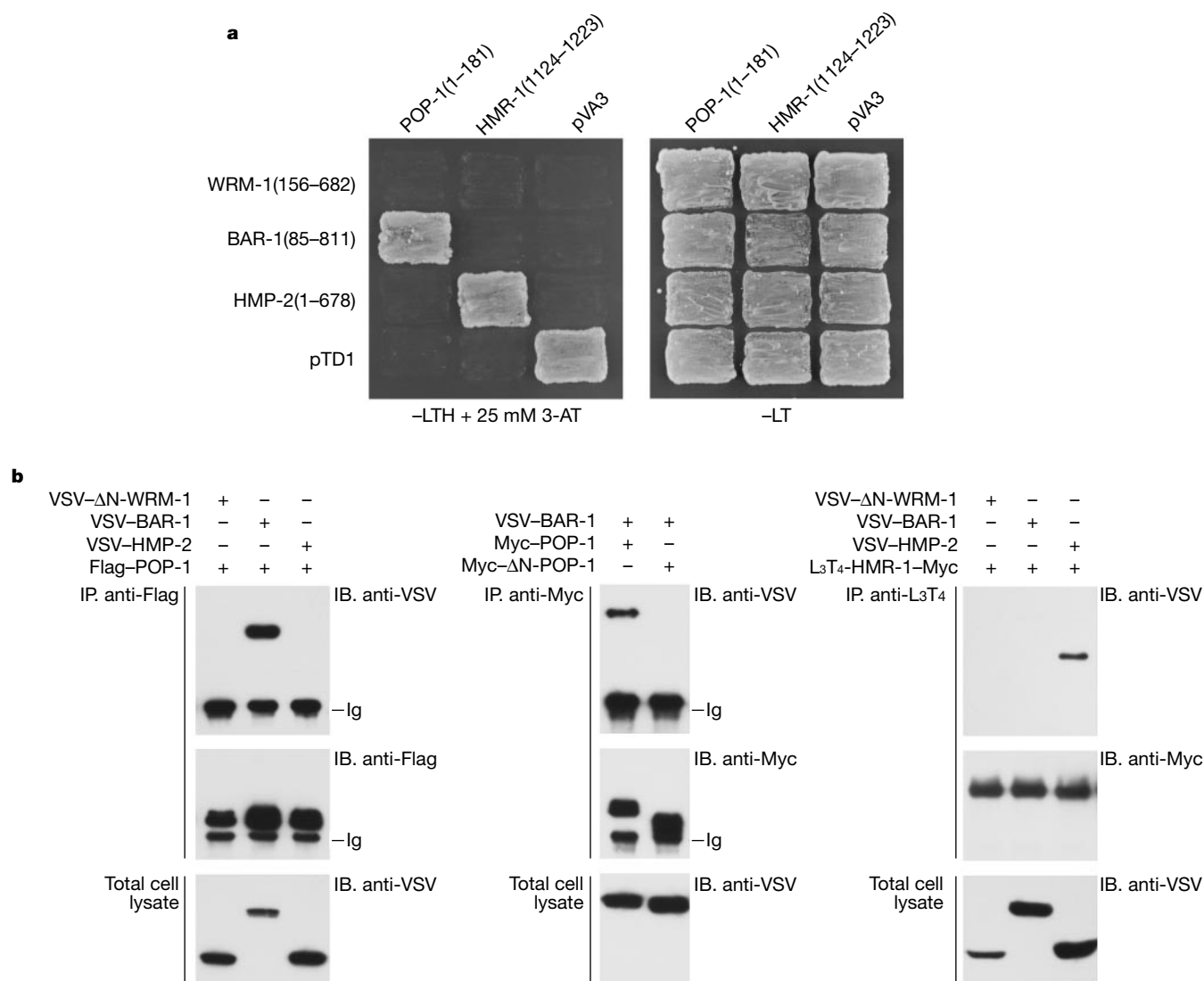


Figure 2 Physical interactions between the three *C. elegans* β-catenins and POP-1 and HMR-1. **a**, Yeast two-hybrid assay showing the specific interactions between POP-1 and BAR-1 and HMR-1 and HMP-2. The WRM-1–Gal4BD fusion does not interact with POP-1 or HMR-1, but does interact with other proteins (data not shown). **b**, Expression vectors were transfected into 293T cells as indicated. VSV-tagged BAR-1 is specifically

co-immunoprecipitated (IP) with Flag-tagged POP-1, as is shown in the anti-VSV immunoblot (IB) (left). This interaction requires the N terminus of POP-1 (middle). Only VSV-tagged HMP-2 is precipitated with the intracellular β-catenin interaction domain of HMR-1 fused to L3T4.

In flies and vertebrates, Armadillo/ β -catenin has a dual function in adhesion and Wg/Wnt signalling. Indeed, Armadillo was first identified for its role in Wg signalling, whereas β -catenin and the related protein Plakoglobin were initially discovered as adhesion molecules^{4,5}. In *C. elegans*, three highly divergent β -catenin homologues have been identified⁶. Of these, HMP-2 is most similar to vertebrate and fly β -catenin⁷. Mutations in *hmp-2* cause defects in hypodermal enclosure and body elongation, indicating that it may be involved in cellular adhesion. Indeed, HMP-2 colocalizes with the cadherin HMR-1 and the α -catenin HMP-1 in adherens junctions⁷. BAR-1 shows a slightly lower overall similarity to β -catenin and is required for several patterning events during larval development^{8,9}. The third homologue, WRM-1, shows the lowest similarity to β -catenin¹. WRM-1 is involved in a divergent Wnt pathway, where it opposes rather than mediates signalling by the Tcf-like transcription factor POP-1 (ref. 1). Instead of binding POP-1 directly, WRM-1 binds and activates a kinase, LIT-1, which in turn phosphorylates POP-1 (refs 2, 3). The activities of WRM-1 and LIT-1 ensure that POP-1, which probably functions as a repressor in this pathway^{10–12}, is asymmetrically distributed between the daughter cells of many embryonic and larval cell divisions, allowing the specification of different cell fates.

POP-1 contains a single Tcf-like HMG-box DNA-binding domain¹⁰ and an acidic region within its amino terminus that has limited sequence similarity to the β -catenin interaction domain of other Tcf members¹³. The HMG domain is highly conserved within the Tcf family, with 92% sequence identity between, for example, human TCF-1 and *Drosophila* dTcf/Pangolin¹³. Despite the fact that POP-1 and TCF-1 share only 54% sequence identity within the HMG-box, we show that POP-1 recognizes the Tcf consensus site. A bacterially expressed POP-1 fragment containing the HMG domain specifically binds the optimal Tcf target site, but not a sequence that lacks the core Tcf target motif (Fig. 1a). Tcf transcription factors bind Armadillo/ β -catenin to activate transcription of target genes^{13–17}. In transient transfections using a previously established β -catenin–Tcf reporter gene assay^{13,18}, we found that co-expression of POP-1 and Armadillo specifically activated transcription of the Tcf reporter, whereas POP-1 alone did not (Fig. 1b). We conclude that POP-1 is a functional Tcf homologue.

Tcf transcription factors interact with Armadillo/ β -catenin by binding to Armadillo repeats 3 to 8 (ref. 13). To determine which of the three *C. elegans* β -catenins binds directly to POP-1, we analysed possible interactions using a yeast two-hybrid assay. POP-1 was fused to the Gal4 DNA-binding domain, and fragments of each *C. elegans* β -catenin homologue containing the 12 Arm repeats were fused with the Gal4 activation domain. A direct physical interaction was only observed between POP-1 and BAR-1 (Fig. 2a). To investigate the specific interaction between POP-1 and BAR-1, we co-expressed Flag-tagged POP-1 with either VSV-tagged Δ N-WRM-1, BAR-1 or HMP-2 in vertebrate cells, and asked which of the three β -catenin homologues immunoprecipitated with POP-1. In this experiment, we used a truncated WRM-1 (Δ N-WRM-1) that lacks the N-terminal LIT-1/Nlk interaction domain², but retains the Arm repeat region. As is shown in Fig. 2b, only BAR-1 was precipitated with POP-1. We found that the interaction between POP-1 and BAR-1 requires the acidic region at the N terminus of POP-1. Removal of the first 44 amino acids of POP-1 (Δ N-POP-1) resulted in the loss of BAR-1 binding (Fig. 2b). These data show that POP-1 specifically interacts with BAR-1.

Next, we tested whether the association between POP-1 and BAR-1 results in the transcriptional activation of a Tcf target gene. Only co-expression of POP-1 and BAR-1 resulted in significant activation of a Tcf reporter gene, whereas co-expression of POP-1 with WRM-1 or HMP-2 had no effect (Fig. 3a). As expected, the transcriptional activation of the Tcf reporter required the BAR-1 interaction domain of POP-1. Thus, co-expression of Δ N-POP-1 with BAR-1 did not activate the Tcf reporter (Fig. 3a). These data indicate that

POP-1 and BAR-1 may function in Wnt target gene activation analogously to Tcf and β -catenin in flies and vertebrates.

To investigate this further, we tested whether POP-1 and BAR-1 are required for the expression of the putative Wnt target gene, *mab-5*. The Hox gene *mab-5* controls the direction of migration of the neuroblast QL and its daughter cells (denoted collectively as QL.d) during larval development^{19,20}. In wild-type animals, the QL daughter cells migrate to positions that are posterior to their point of origin. In *mab-5* loss-of-function mutants, the QL.d cells migrate in the opposite, anterior direction. The expression of *mab-5* in the QL lineage is regulated by a putative Wnt pathway. Mutations in *egl-20/Wnt*^{19,21}, *lin-17/Fz*^{21,22} and *bar-1* (refs 8, 9) all result in anterior migration of the QL.d cells and produce a loss of *mab-5* expression in the QL lineage.

The only existing *pop-1* mutation, *pop-1(zu189)*, produces a maternal-effect lethal phenotype¹⁰ and has no zygotic effect on QL.d migration. Therefore, we could not directly investigate the function of *pop-1* in this pathway by using this mutation. We have previously used overexpression of truncated Tcf mutants that lack the β -catenin interaction domain to produce potent dominant-negative mutants^{13,15}. As is shown in Fig. 3b, overexpression of Δ N-POP-1 strongly inhibited Tcf reporter activation by full-length POP-1 and BAR-1. We generated stable transgenic lines that inducibly express Δ N-POP-1 from a heat-shock promoter. The effect of Δ N-POP-1 expression on QL.d migration was assessed by scoring the final positions of the two QL.pa daughter cells, which can be easily recognized, relative to the positions of the anterior and posterior daughters of the six seam cells, V1 to V6 (ref. 21) (Fig. 4a). Whereas the QL.pa daughter cells were found at a normal position (near V5.a) in non-heat-shocked animals, a 1-h heat shock at hatching resulted in the anterior migration of these cells, with >80% of QL.pa daughters localizing anterior to V2.p. This migration defect is similar to the phenotype of *egl-20/Wnt* and *mab-5* null mutants²¹, and indicates that overexpression of Δ N-POP-1 may abrogate the expression of *mab-5*. To verify this, we used a *mab-5::lacZ* reporter construct that mimics the expression

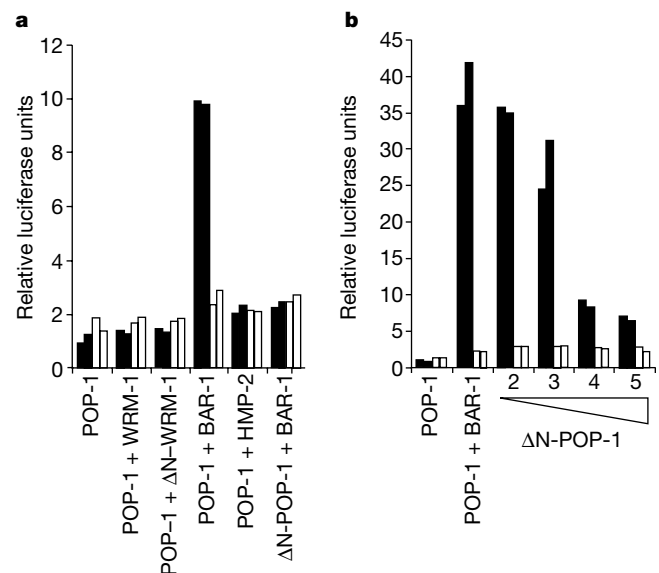


Figure 3 POP-1 and BAR-1 activate transcription of a Tcf reporter gene. **a**, Expression constructs were transfected into IIA.6 cells with the TOP reporter (black bars) or the negative control FOP (white bars) as indicated. Only POP-1 and BAR-1 activate the Tcf reporter. The BAR-1 interaction domain of POP-1 is required for this activity (Δ N-POP-1). **b**, Overexpression of Δ N-POP-1 inhibits activation of the Tcf reporter by full-length POP-1 and BAR-1. Δ N-POP-1 was transfected at increasing concentrations (2–5 μ M) with full-length POP-1 and BAR-1.

of the endogenous *mab-5* gene²⁰. We found that 26/26 non-heat-shocked animals expressed the *mab-5::lacZ* reporter in the QL daughter cells, whereas 34/34 heat-shocked animals did not (Fig. 4b), indicating that overexpression of Δ N-POP-1 inhibits expression of *mab-5*. Consistent with these observations, RNA-mediated interference of *pop-1* zygotic function (see Methods) also resulted in the anterior migration of the QL.d cells and the loss of *mab-5* expression in the QL lineage (M.H., in preparation). We found that the other two β -catenin homologues, WRM-1 and HMP-2, are not required for QL.d migration. RNA-mediated interference of either *hmp-2* ($n = 21$) or *wrm-1* zygotic functions ($n = 33$), did not result in QL.d migration defects. We conclude that POP-1 and BAR-1 function in a canonical Wnt pathway that regulates the expression of target genes such as *mab-5*.

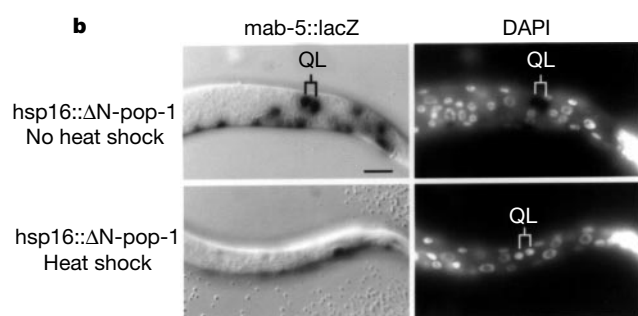
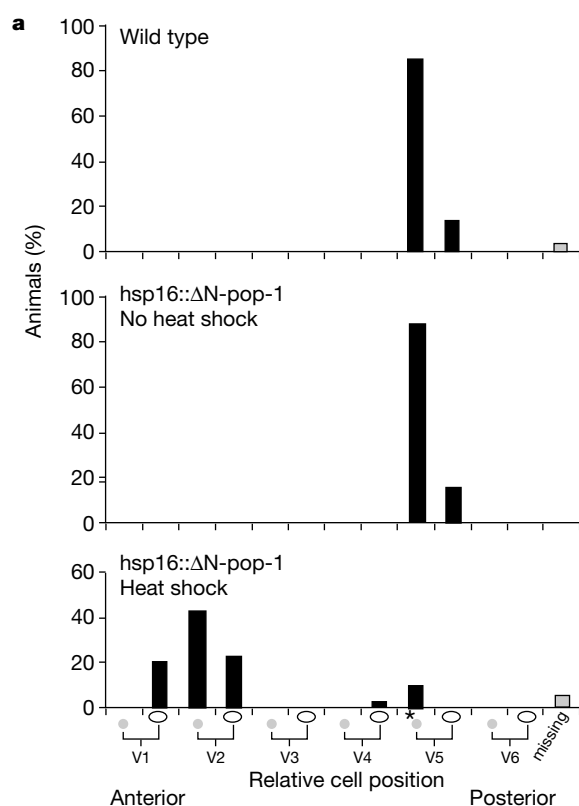


Figure 4 Expression of Δ N-POP-1 inhibits *mab-5* expression. **a**, Expression of Δ N-POP-1 caused anterior migration of the QL.d cells. Final positions of the QL.pa daughters were scored²¹ in wild-type ($n = 32$), non-heat-shocked ($n = 29$) and heat-shocked ($n = 45$) animals containing *huls4*. Asterisk above V5.a marks the birth place of QL. Black bars indicate the proportion of cells at each position. The grey bar indicates the proportion of missing cells. **b**, Expression of Δ N-POP-1 blocked the expression of a *mab-5::lacZ* reporter construct in the QL.d cells. Whole-mount larvae, 3–5 h after hatching, were fixed and stained for β -galactosidase expression²⁰ and the DNA stain DAPI.

As discussed above, vertebrate and fly Armadillo/ β -catenin also functions as an adhesion molecule by interacting with the cytoplasmic tail of cadherins. A single classical cadherin, HMR-1, has been identified in *C. elegans*⁷ and extensive database searches of the now virtually complete genome sequence have revealed no other candidates. We investigated which of the three *C. elegans* β -catenins bind HMR-1. The β -catenin interaction domain of HMR-1 (ref. 7) was fused to the Gal4 DNA-binding domain (BD) and tested for direct physical interaction with the three *C. elegans* β -catenins in a yeast two-hybrid assay. HMR-1 interacted only with HMP-2 (Fig. 2a). To investigate this further, we fused the extracellular part of mouse CD4 (L3T4) with the transmembrane and intracellular domains of HMR-1 to localize HMR-1 at the plasma membrane. We co-expressed the HMR-1 fusion protein with VSV-tagged Δ N-WRM-1, BAR-1 or HMP-2 in vertebrate cells and investigated which of the *C. elegans* β -catenin homologues interacted with the intracellular domain of HMR-1. Only HMP-2 immunoprecipitated

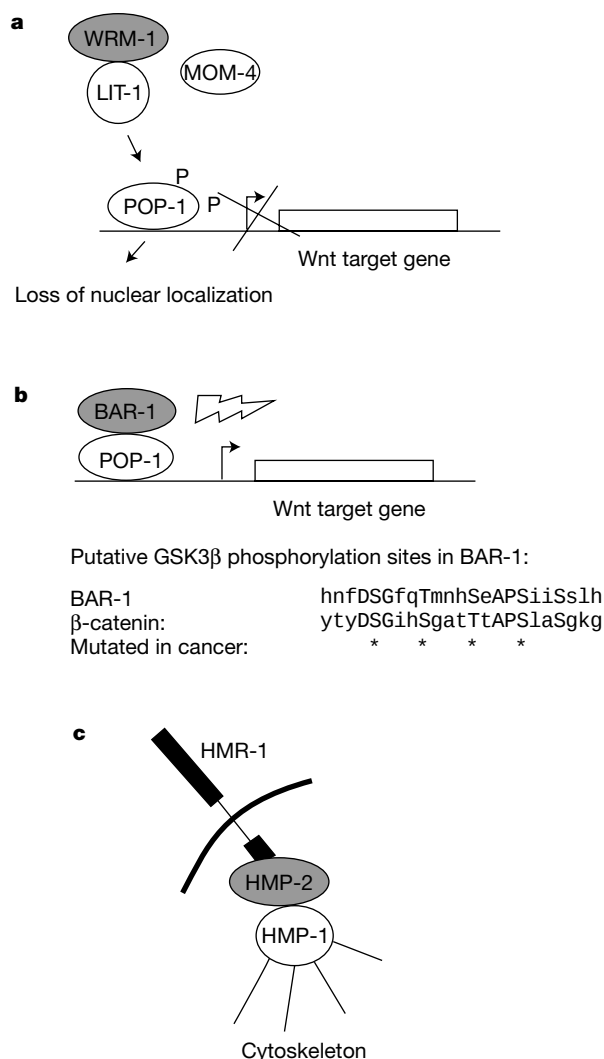


Figure 5 The signalling and adhesion functions of Armadillo/ β -catenin are distributed over three different β -catenin homologues in *C. elegans*. **a**, WRM-1 acts in a divergent Wnt pathway. WRM-1 binds to and activates the kinase LIT-1, which in turn phosphorylates POP-1. Phosphorylation of POP-1 may result in its loss of nuclear localization^{2,24}. **b**, BAR-1 is part of a canonical Wnt pathway. BAR-1 and POP-1 have a function in Wnt target gene activation that is analogous to the function of β -catenin and Tcf in flies and vertebrates. BAR-1 is the only *C. elegans* β -catenin that harbours a set of putative GSK3 β phosphorylation sites in its N-terminal sequence. **c**, HMP-2 is the only *C. elegans* β -catenin that interacts with the cadherin HMR-1.

with HMR-1 (Fig. 2b), again showing a specific interaction between these two proteins. Thus the cadherin HMR-1 interacts specifically with HMP-2, but not with WRM-1 or BAR-1.

We propose that the signalling and adhesion functions of Armadillo/ β -catenin have been distributed between separate β -catenin homologues in *C. elegans* (Fig. 5). Two β -catenins function in Wnt signalling. WRM-1 is part of a divergent Wnt pathway that, in collaboration with a mitogen-activated protein (MAP) kinase pathway, mediates the asymmetric distribution of POP-1 between daughter cells of many anterior/posterior cell divisions^{1–3,11,12,23} (reviewed in ref. 24). POP-1 probably acts as a repressor in this pathway^{10–12} and differences in POP-1 expression levels between daughter cells may allow the specification of different fates. We show that BAR-1 is part of a Wnt pathway that is similar to that in flies and vertebrates. BAR-1 can directly associate with POP-1 to activate a Tcf reporter. In addition, we show that BAR-1 and POP-1 are required for the expression of the endogenous Wnt target gene *mab-5*. Of the three *C. elegans* β -catenins, only BAR-1 contains a set of four conserved GSK3 β phosphorylation sites⁸ (Fig. 5b). These sites are essential for the regulation of β -catenin by the Wnt pathway and are frequently mutated in cancers^{25,26}. *C. elegans* contains a single putative APC homologue, APR-1¹. In vertebrates, APC forms a complex with GSK3 β , Axin and β -catenin to downregulate β -catenin levels in the absence of Wnt signalling^{4,5}. We find that none of the *C. elegans* β -catenins bind APR-1 in a yeast two-hybrid assay (data not shown). Together with the observation that the *C. elegans* genome does not contain a clear Axin homologue⁶, this indicates that BAR-1 may be regulated differently.

HMP-2 is the only *C. elegans* β -catenin that interacts with the single classical cadherin, HMR-1. This interaction agrees with the *hmp-2* mutant phenotype and with the colocalization of HMP-2 with HMR-1 and the α -catenin HMP-1 in adherens junctions⁷. We conclude that HMP-2 functions specifically in adhesion (Fig. 5c). Vertebrates express a second β -catenin-like protein, Plakoglobin, which is part of the desmosomal adhesion complex. It is unclear whether Plakoglobin functions specifically in adhesion or also has a role in Wnt signalling²⁷. The functions of Armadillo and β -catenin in Wnt signalling and adhesion are physically separable^{28,29}. It is unclear whether Armadillo/ β -catenin in adherens junctions may directly or indirectly affect the cytoplasmic signalling pool. Mutations in cadherins have been identified in many epithelial tumours. An attractive explanation for the oncogenic potential of cadherin mutations is the release of β -catenin from adherens junctions, which in turn can interact with Tcf transcription factors to activate the expression of Wnt target genes³⁰. Our data indicate that, at least in *C. elegans*, the adhesion and signalling pools of β -catenin are separate entities. □

Methods

Cloning of complementary DNAs

We cloned full-length cDNA of *pop-1* (GenBank accession no. U37532), *bar-1* (AF063646), *wrm-1* (AF013951) and *hmp-2* (AF016853) and parts of the *hmr-1* (AF016854) cDNA by polymerase chain reaction on total *C. elegans* cDNA, and checked them by sequencing. The cDNAs were cloned in-frame with N-terminal Flag, Myc or VSV tags in pCDNA3 (Invitrogen). Δ N-WRM-1 has an N-terminal truncation of the first 134 amino acids. A *hmr-1* fragment encoding amino-acids 904–1223 was cloned in-frame with the extra-cellular part of L3T4 and a carboxy-terminal Myc-tag.

Gel-shift experiments

A POP-1(180–280) fragment containing the HMG-box DNA-binding domain was purified as a His-tagged protein. Gel shifts were performed as described¹⁸.

Interaction studies

We performed two-hybrid experiments as in ref. 15. pVA3 encodes a murine p53–Gal4 binding domain fusion in pGBT9, and pTD1 contains SV40 large T in pGADGH (Clontech). Expression vectors were transfected into 293T cells using FuGENE transfection reagent (Boehringer Mannheim). Cells were lysed in Triton X-100 lysis buffer²⁶ and immunoprecipitations were performed with anti-Flag M2 (Sigma), anti-Myc (9E10), anti-VSV-G (P5D4) or anti-L3T4 (GK1.5) (Pharmingen) and protein A/G beads (Santa Cruz).

Luciferase reporter assays

Luciferase reporter assays were performed as described^{13,18}. Briefly, we transfected 2×10^6 IIAL6 cells with 1 μ g of the luciferase reporter pTKTOP or the negative control pTKFOP, 1 μ g of a POP-1 expression construct, 5 or 10 μ g of the different β -catenin expression constructs and 50 ng of the internal transfection control pTKRenilla. Luciferase assays were as recommended by the manufacturer (Promega). Luciferase measurements were normalized for transfection efficiency using the renilla control.

Transgenic animals, RNA-mediated interference and scoring QL.d cell migrations

We cloned a truncated *pop-1* fragment encoding Δ N-POP-1(45–438) into the heat-shock promoter vector pPD49.78 and injected it with the *rol-6(su1006)* marker plasmid pRF4. Animals containing the integrated transgene *hul54* were heat-shocked at hatching for 1 h at 33 °C. RNA-mediated interference (RNAi) was performed as described¹. To determine the zygotic phenotype of maternal effect lethal genes such as *pop-1* and *wrm-1*, we injected double-stranded RNA into an RNAi resistant mutant and analysed the phenotypes of the cross-progeny (M.H., in preparation). The QL.pa daughter cells were scored in the late L1 stage as described²¹. *mab-5* expression was assayed using the *mab-5::lacZ* reporter transgene *muls2* (ref. 20).

Received 14 March; accepted 17 May 2000.

1. Rocheleau, C. E. *et al.* Wnt signaling and an APC-related gene specify endoderm in early *C. elegans* embryos. *Cell* **90**, 707–716 (1997).
2. Rocheleau, C. E. *et al.* WRM-1 activates the LIT-1 protein kinase to transduce anterior/posterior polarity signals in *C. elegans*. *Cell* **97**, 717–726 (1999).
3. Shin, T. H. *et al.* MOM-4, a MAP kinase kinase kinase-related protein, activates WRM-1/LIT-1 kinase to transduce anterior/posterior polarity signals in *C. elegans*. *Mol. Cell* **4**, 275–280 (1999).
4. Miller, J. R. & Moon, R. T. Signal transduction through β -catenin and the specification of cell fate during embryogenesis. *Genes Dev.* **10**, 2527–2539 (1996).
5. Cadigan, K. M. & Nusse, R. Wnt signaling: a common theme in animal development. *Genes Dev.* **11**, 3286–3305 (1997).
6. Ruvkun, G. & Hobert, O. The taxonomy of developmental control in *Caenorhabditis elegans*. *Science* **282**, 2033–2041 (1998).
7. Costa, M. *et al.* A putative catenin-cadherin system mediates morphogenesis of the *Caenorhabditis elegans* embryo. *J. Cell Biol.* **141**, 297–308 (1998).
8. Eisenmann, D. M., Maloof, J. N., Simske, J. S., Kenyon, C. & Kim, S. K. The β -catenin homolog BAR-1 and LET-60 Ras co-ordinately regulate the Hox gene *lin-39* during *Caenorhabditis elegans* vulval development. *Development* **125**, 3667–3680 (1998).
9. Maloof, J. N., Whangbo, J., Harris, J. M., Jongeward, G. D. & Kenyon, C. A Wnt signaling pathway controls Hox gene expression and neuroblast migration in *C. elegans*. *Development* **126**, 37–49 (1999).
10. Lin, R., Thompson, S. & Priess, J. R. *pop-1* encodes an HMG box protein required for the specification of a mesoderm precursor in early *C. elegans* embryos. *Cell* **83**, 599–609 (1995).
11. Lin, R., Hill, R. J. & Priess, J. R. POP-1 and anterior-posterior fate decisions in *C. elegans* embryos. *Cell* **92**, 229–239 (1998).
12. Thorpe, C. J., Schlessinger, A., Carter, J. C. & Bowerman, B. Wnt signaling polarises an early *C. elegans* blastomere to distinguish endoderm from mesoderm. *Cell* **90**, 695–705 (1997).
13. van de Wetering, M. *et al.* Armadillo coactivates transcription driven by the product of the *Drosophila* segment polarity gene *dTCF*. *Cell* **88**, 789–799 (1997).
14. Behrens, J. *et al.* Functional interaction of β -catenin with the transcription factor LEF-1. *Nature* **382**, 638–642 (1996).
15. Molenaar, M. *et al.* XTcf-3 transcription factor mediates β -catenin-induced axis formation in *Xenopus* embryos. *Cell* **86**, 391–399 (1996).
16. Riese, J. *et al.* LEF-1, a nuclear factor co-ordinating signaling inputs from *wingless* and *decapentaplegic*. *Cell* **88**, 777–787 (1997).
17. Brunner, E., Peter, O., Schweizer, L. & Basler, K. *pangolin* encodes a Lef-1 homologue that acts downstream of Armadillo to transduce the Wingless signal in *Drosophila*. *Nature* **385**, 829–833 (1997).
18. Korinek, V. *et al.* Constitutive transcriptional activation by a β -catenin-Tcf complex in APC^{−/−} colon carcinoma. *Science* **275**, 1784–1787 (1997).
19. Kenyon, C. A gene involved in the development of the posterior body region of *C. elegans*. *Cell* **46**, 477–487 (1986).
20. Salser, S. & Kenyon, C. Activation of a *C. elegans Antennapedia* homologue in migrating cells controls their direction of migration. *Nature* **355**, 255–258 (1992).
21. Harris, J., Honigberg, L., Robinson, N. & Kenyon, C. Neuronal cell migration in *C. elegans*: regulation of Hox gene expression and cell position. *Development* **122**, 3117–3131 (1996).
22. Sawa, H., Lobel, L. & Horvitz, H. R. The *Caenorhabditis elegans* gene *lin-17*, which is required for certain asymmetric cell divisions, encodes a putative seven-transmembrane protein similar to the *Drosophila* Frizzled protein. *Genes Dev.* **10**, 2189–2197 (1996).
23. Meneghini, M. D. *et al.* MAP kinase and Wnt pathways converge to downregulate an HMG-domain repressor in *Caenorhabditis elegans*. *Nature* **399**, 793–797 (1999).
24. Thorpe, C. J., Schlessinger, A. & Bowerman, B. Wnt signalling in *Caenorhabditis elegans*: regulating repressors and polarising the cytoskeleton. *Trends Cell Biol.* **10**, 10–17 (2000).
25. Morin, P. J. *et al.* Activation of β -catenin-Tcf signaling in colon cancer by mutation in β -catenin or APC. *Science* **275**, 1787–1790 (1997).
26. Rubinfeld, B. *et al.* Stabilisation of β -catenin by genetic defects in melanoma cell lines. *Science* **275**, 1790–1792 (1997).
27. Simcha, I. *et al.* Differential nuclear translocation and transactivation potential of β -catenin and Plakoglobin. *J. Cell Biol.* **141**, 1433–1448 (1998).
28. Orsolic, S. & Peifer, M. An *in vivo* structure-function study of Armadillo, the β -catenin homologue, reveals both separate and overlapping regions of the protein required for cell adhesion and for Wingless signaling. *J. Cell Biol.* **134**, 1283–1300 (1996).
29. Fagotto, F., Funayama, N., Glück, U. & Gumbiner, B. M. Binding to cadherins antagonises the signaling activity of β -catenin during axis formation in *Xenopus*. *J. Cell Biol.* **132**, 1105–1114 (1996).

30. Christofori, G. & Semb, H. The role of the cell-adhesion molecule E-cadherin as a tumour-suppressor gene. *Trends Biochem. Sci.* **24**, 73–76 (1999).

Acknowledgements

We thank L. Meysaard for critically reading the manuscript and members of the Clevers laboratory for helpful discussions; C. Kenyon for the *mab-5* reporter *mulS2*; Q. Ch'ng and C. Kenyon for tips on staining *mulS2* animals for β -galactosidase expression; and A. Fire for pPD49.78. This work was supported in part by an NIH grant to M.H. and PIONEER and Program grants from NWO Medische Wetenschappen to H.C.

Correspondence and requests for materials should be addressed to H.K. (e-mail: R.Korswagen@lab.azn.nl).

Genomic analysis of metastasis reveals an essential role for RhoC

Edwin A. Clark^{*†}, Todd R. Golub^{‡§}, Eric S. Lander^{‡||} & Richard O. Hynes^{*||}

^{*} Howard Hughes Medical Institute, Centre for Cancer Research, and

^{||} Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

[‡] Whitehead Institute/MIT Centre for Genome Research, 9 Cambridge Centre, Cambridge, Massachusetts 02142, USA

[§] Dana-Farber Cancer Institute, 44 Binney Street, Boston, Massachusetts 02115, USA

The most damaging change during cancer progression is the switch from a locally growing tumour to a metastatic killer. This switch is believed to involve numerous alterations that allow tumour cells to complete the complex series of events needed for metastasis¹. Relatively few genes have been implicated in these events^{2–5}. Here we use an *in vivo* selection scheme to select highly metastatic melanoma cells. By analysing these cells on DNA arrays, we define a pattern of gene expression that correlates with progression to a metastatic phenotype. In particular, we show enhanced expression of several genes involved in extracellular matrix assembly and of a second set of genes that regulate, either directly or indirectly, the actin-based cytoskeleton. One of these, the small GTPase RhoC, enhances metastasis when overexpressed, whereas a dominant-negative Rho inhibits metastasis. Analysis of the phenotype of cells expressing dominant-negative Rho or RhoC indicates that RhoC is important in tumour cell invasion. The genomic approach allows us to identify families of genes involved in a process, not just single genes, and can indicate which molecular and cellular events might be important in complex biological processes such as metastasis.

To provide insight into the pattern of gene expression that allows tumours to metastasize, we compared the gene expression profile of melanoma variants with low or high metastatic potential. As shown in Fig. 1, the system involves the *in vivo* selection of highly metastatic melanoma cells from a population of poorly metastatic tumour cells⁶. When nude mice were injected intravenously with amelanotic human A375P tumour cells, relatively few pulmonary metastases were observed (Fig. 2a). When these rare metastases were dissected free from the lungs and the cells grown in tissue culture, however, the resulting cells showed enhanced metastatic capacity, confirming that highly metastatic cells can be selected from a heterogeneous population of poorly metastatic tumour cells⁷. Furthermore, if successive metastases (designated M1 and M2) were isolated, expanded in tissue culture, and re-introduced into

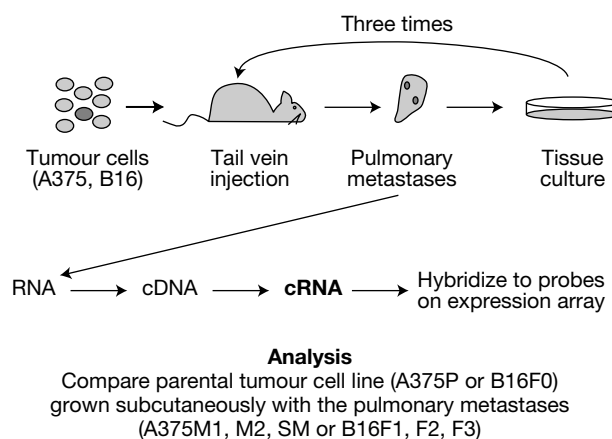


Figure 1 *In vivo* selection scheme. Poorly metastatic melanoma cell lines (human A375P or mouse B16F0) were injected intravenously into the tail veins of host mice and pulmonary metastases were isolated. Either these metastases were minced and grown in tissue culture (to be injected into additional host mice) or RNA was extracted to prepare the labelled cRNA used to hybridize to the oligonucleotide arrays. The procedure to select for highly metastatic tumour cells was repeated two (A375) or three (B16) times. A375SM cells were previously derived in a similar manner¹¹.

host mice as shown in Fig. 1, significantly more pulmonary metastases were observed (Fig. 2b). When mouse B16F0 melanoma cells were subjected to this same *in vivo* selection scheme, highly metastatic pulmonary tumours (designated F1, F2 and F3) were isolated, as previously described for this cell line⁶. When the poorly metastatic A375P or B16F0 and the *in vivo*-selected metastatic A375 or B16 cells were grown as subcutaneous tumours, there was no observable difference in tumour size (see Supplementary Information), indicating that we had selected for a difference in metastatic, but not tumorigenic, properties of the melanomas. These results support the hypothesis that specific gene products can regulate metastasis without altering the growth properties of a tumour⁸. Therefore, we sought to identify metastasis-specific genes using a functional genomics approach.

RNAs extracted from these pulmonary metastases and from the parental A375P and B16F0 lines grown as subcutaneous tumours were used to prepare complementary RNAs (cRNAs), which were hybridized to oligonucleotide microarrays (human: 7,070 genes; mouse: 6,347 genes, with around 50% overlap in the genes represented) to determine the array of differentially expressed genes (Fig. 1). The entire data set is available at our web site at <http://www.genome.wi.mit.edu/MPR> and in Supplementary Information. Table 1 lists those genes expressed at consistently higher levels in pulmonary metastases derived from the A375P line (M1, M2 and SM) and the mouse B16F0 line (F1, F2 and F3). To ensure that the enhanced expression of these genes in the pulmonary metastases was not due solely to the influence of the microenvironment in which the metastatic cells were growing, we also grew metastatic A375SM cells subcutaneously and compared their expression profile with that of subcutaneous A375P tumours. We found that 15 of the 16 genes continued to show enhanced expression when metastatic A375 cells were grown as a subcutaneous tumour (see Supplementary Information), indicating that the expression of these genes is intrinsic to the metastatic cells. Note, however, that the tumour microenvironment may help to regulate the absolute level of gene expression.

As the set of genes represented on the human and mouse arrays partially overlapped, some signals appeared in both species (Table 1). Three genes, fibronectin, RhoC and thymosin β 4, were expressed at higher levels (≥ 2.5 -fold) in all three metastases selected from both the human A375 and mouse B16 cell lines. Enhanced expression of these three genes in the pulmonary metastases was

[†] Present address: Millennium Predictive Medicine, One Kendall Square, Cambridge, Massachusetts 02139, USA.

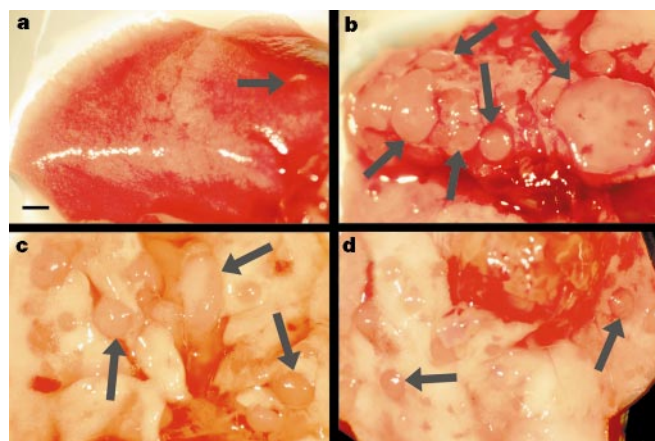


Figure 2 Pulmonary metastases in lungs of mice injected with tumour cell lines. **a**, A375P cells; **b**, A375M; **c**, A375P-RhoC; **d**, A375M-dnRho. Arrows indicate representative metastatic nodules. Scale bar, 1 mm.

confirmed by RNase protection (see Supplementary Information). Note that this assay will detect only human RNAs (from the tumour cells) and not mouse RNAs from stromal or vascular cells. This confirms that changes in expression of these three genes occur in the tumour cells rather than in host tissue. This is also probably true for all the oligonucleotide array data given the sequence divergence between species (Table 1).

Fibronectin is an extracellular glycoprotein that serves as a ligand for the integrin family of cell adhesion receptors and regulates cytoskeletal organization. Fibronectin expression has been linked with tumorigenesis⁹ and metastasis¹⁰, although these studies are

only correlative. Peptides that mimic the cell adhesive region of fibronectin are, however, known to inhibit metastasis¹¹, which may indicate that tumour cells must interact with molecules such as fibronectin to metastasize. RhoC is a member of the Rho GTPase family that can regulate many cellular functions, most notably cytoskeletal organization, in response to extracellular factors¹². Enhanced expression of RhoC has been reported to correlate with the progression of pancreatic adenocarcinomas to a metastatic phenotype¹³. Thymosin β 4 is an actin-sequestering protein that regulates actin polymerization; its expression in renal tumours has been correlated with malignancy¹⁴. Expression of two other family members, thymosin β 10 and thymosin β 15, also correlates with metastasis^{4,15}. Other regulators of the cytoskeleton also appear on the list, including α -catenin and expressed sequence tags (ESTs) for α -actinin 1 and α -centractin. The altered expression of so many genes whose products regulate the actin cytoskeleton either directly or indirectly indicates that cytoskeletal organization may be important in tumour metastasis.

Also prominent on the list in Table 1 are several genes that encode extracellular matrix (ECM) proteins, as well as molecules that regulate their assembly. In addition to fibronectin, two collagen subunits, α 2(I) and α 1(III), the matrix Gla protein, fibromodulin and biglycan also are expressed at higher levels in the metastatic melanomas. Previous studies of matrix Gla protein have shown that it is overexpressed in breast carcinoma cell lines relative to normal breast epithelial cells¹⁶ and collagen expression has been correlated with the invasive potential of ocular melanomas¹⁰, but expression of the small interstitial proteoglycans biglycan and fibromodulin (which regulate collagen fibril formation^{17,18}) has yet to be linked to tumour progression. These findings support hypotheses that enhanced expression of ECM proteins may promote tumour cell survival or angiogenesis¹⁹.

Table 1 Enhanced gene expression in metastatic melanomas

Gene name	human A375						mouse B16					
	Human acc. number	Ch. no.	P	M1	M2	SM	Mouse acc. number	F0	F1	F2	F3	Nuc. ident.
Fibronectin	X02761	2	1	10.1	3.2	4.0	M18194	A	2.8	2.8	2.8	93%
RhoC	L25081	1	A	4.7	3.1	2.8	X80638	A	2.9	4.9	2.5	91%
Thymosin β 4	M17733	X	1	3.3	3.6	3.5	W41883	1	4.1	3.5	3.5	92%
t-PA	K03021	8	A	5.2	9.6	5.2	J03520	A	A	A	A	81%
Angiopoietin1	D13628	8	1	4.3	9.4	3.3	U83509			*		
IEX-1/Glu96	S81914	6	1	9.1	3.3	4.5	X67644	1	0.4	0.6	0.5	83%
RTP/NDR1	D87953	8	1	8.6	5.4	4.7	U60593	1	A	0.7	1.5	86%
Fibromodulin	U05291	1	A	8.3	4.7	8.2	X94998	1	2.0	2.0	1.1	80%
Hsp70	M11717	1	1	7.8	4.2	5.0	M20567	1	2.1	1.8	1.8	80%
IL13 Rec., α 2	U70981	X	1	7.6	2.9	3.1	U65747					
Sec61 β	L25085	9	1	3.8	4.7	3.3				†		
snRNP, poly. pep. C	HG1322	9	1	3.8	5.3	3.2				†		
Collagen α 2	Z74616	7	A	2.5	3.6	3.6	X58251	A	3.1	2.3	3.7	86%
UBE21	U45328	16	1	3.6	3.4	3.4				†		
KIAA0156	D63879	16	1	3.6	3.4	3.4				†		
TGF β superfamily	AB000584	19	1	3.4	3.4	3.0				†		
Surfactant protein C	J03890		*				M38314	A	32	12	16	
Lysozyme M			†				M21050	A	20	10	22	
Matrix Gla prot	X53331	12	1	3.2	4.4	1.1	D00613	1	12	11	5.4	81%
Tsa-1			†				U47737	A	9.7	6.1	7.2	
Collagen III α 1	X06700	2	A	A	A	A	X52046	A	8.2	5.6	5.5	89%
Biglycan	J04599	X	A	A	3.7	A	L20276	A	3.8	4.4	6.9	87%
α -catenin	U03100	5	1.0	1.3	1.0	1.9	X59990	1	3.4	3.0	5.7	91%
Valosin-cont. prot.	AC004472		*				Z14044	1	3.0	3.9	5.9	
ERK-1	X60188	16	A	A	A	A	Z14249	1	2.6	2.6	3.0	85%
α -actinin 1							AA068062	1	3.6	3.3	7.3	
calmodulin							AA103356	A	4.8	6.7	5.5	
EIF4 γ							AA002277	A	4.7	3.2	2.6	
α -centractin							W48490	1	2.9	3.8	3.6	
IQGAP1							AA118739	A	3.6	3.5	3.2	
cathepsin s							W13263	A	2.8	2.8	3.1	
EF2							W90866	1	2.6	2.5	2.9	

Genes whose expression is consistently enhanced >2.5-fold in pulmonary metastases (M1, M2, SM, F1, F2 or F3) compared to poorly metastatic cells (P or F0) grown as subcutaneous tumours. The values for P and F0 are the average of two experiments performed with subcutaneous tumours from two mice injected with A375P or B16F0 cells. Data are presented as fold expression compared with the poorly metastatic tumours. When expression was below baseline, the expression was marked as absent (A) and was arbitrarily set at 2.0. Mouse expressed sequence tags (ESTs) are noted in italics and are named according to the gene to which they show the greatest sequence similarity. Ch. no., human chromosome where the gene resides. Nuc. ident., percentage of nucleotides identical between the human and mouse homologues, as determined by BLAST search. The accession number is the GenBank entry from which the oligonucleotide probe sequences were drawn.

* Mouse or human gene homologue exists in the UNIGENE database but was not part of the oligonucleotide probe set.

† No gene homologue was found in the UNIGENE database.

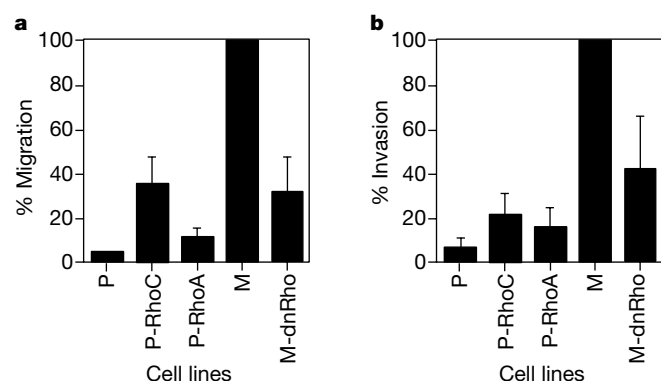


Figure 3 RhoC regulates melanoma cell chemotaxis and invasion. **a**, Poorly (P) or highly metastatic (M) A375 cells expressing rhoC, rhoA or dominant-negative (dn) rho were induced to migrate towards serum for 16 h. Over 25% of the plated A375M cells migrated within this time. Each bar represents the mean $\pm$ s.e.m. of four experiments done in duplicate. **b**, The cell lines described above were induced to invade matrigel-coated membranes for 48 h. Over 30% of the plated A375M cells invaded within this time. Each bar represents the mean $\pm$ s.e.m. of three experiments done in duplicate.

Several genes identified in other studies are conspicuous by their absence from our list. Metastasis suppressor genes, such as nm23, KiSS1 and CD82, can inhibit tumour metastasis². In our study all three of these genes were absent in both the parental A375 tumours and in the metastases (see Supplementary Information), indicating that, although expression of these genes may inhibit metastasis, lack of their expression is not sufficient for metastasis. Other genes not found in Table 1 but whose expression correlated with melanoma metastasis in previous studies include the Met tyrosine kinase receptor, matrix metalloproteinases (MMPs) such as MMP2, and the β 3-integrin subunit^{20–22}. In the B16 tumours, Met expression was higher in two of the three metastases but its expression was not detected in any of the A375 tumours, indicating that its expression may not be essential for these tumours to metastasize. Expression of MMP2 and of the β 3-integrin subunit was not significantly higher in any of the three metastases (see Supplementary Information), but their expression in both the parental and metastatic tumours may be sufficient to allow the tumour cells to metastasize.

Having uncovered 32 genes and ESTs whose expression patterns correlate with metastasis, we wished to investigate the function of one of these genes in this process. Because of its elevated expression in metastases derived from both tumour cell lines, RhoC was chosen to test the hypothesis that these expression studies identify genes essential for metastasis. The full-length human RhoC gene was cloned, subcloned into a retroviral vector and introduced into a retroviral packaging cell line. We used retroviral particles to infect the poorly metastatic A375P cells, and selected cells expressing high levels of RhoC by fluorescence-activated cell sorting (FACS). These cells, designated A375P-RhoC (expressing RhoC at 20 times the level expressed in A375M cells), were subjected to the experimental metastasis assay. As seen in Fig. 2c and Table 2, overexpression of RhoC markedly enhanced metastasis in this system.

Next we tested whether we could inhibit metastasis by expressing a known dominant-inhibitory Rho mutant (N19Rho)²³ in the highly metastatic A375M cells (a pool of M1, M2 and SM cells). This mutant is analogous to the N17Ras mutant that blocks Ras signalling²⁴. Ras dominant-negatives are actually antagonists of the guanine-nucleotide exchange factors (GNEFs) for Ras, rather than of Ras itself²³ and it is believed that dominant-negative RhoA antagonizes Rho GNEFs, thereby inhibiting RhoC. Expression of N19RhoA in the A375M cells markedly suppressed the generation of metastases when these cells were subjected to the experimental metastasis assay (Fig. 2d and Table 2), indicating that Rho activity may be necessary, and RhoC sufficient, for metastasis in these melanoma lines.

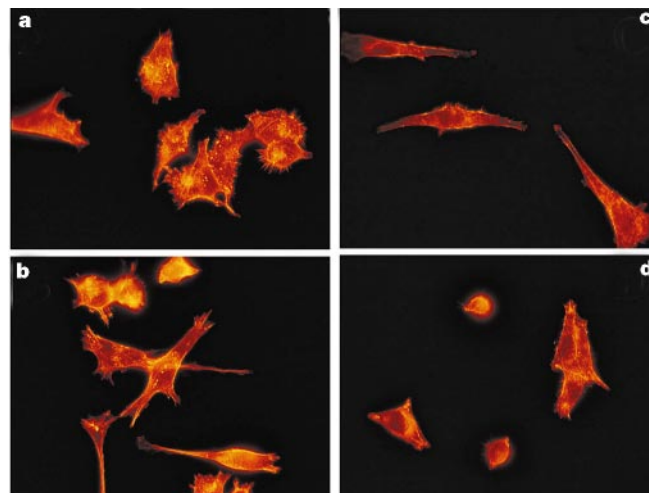


Figure 4 Metastatic capacity of A375 melanoma cells correlates with cell morphology. Poorly (**a**) or highly metastatic (**c**) A375 cells expressing rhoC (**b**) or dominant-negative rho (**d**) were plated on glass coverslips for 16 h at 37 °C, then fixed and stained with phalloidin to detect F-actin.

Having established a causal role for RhoC in metastasis, we set out to investigate how RhoC might regulate the ability of tumour cells to metastasize. Tumour cells must complete a complex series of steps to metastasize, one of the most basic of which is cell growth. Rho GTPases affect several aspects of growth control¹², so it was possible that RhoC might control tumour metastasis by regulating cell proliferation. We assayed the A375P, A375P-RhoC, A375M and A375M-dnRho cells for *in vitro* proliferation and *in vivo* tumorigenesis (see Supplementary Information). Proliferation in either assay was not significantly changed by altering RhoC expression or Rho activity, indicating that RhoC may regulate metastasis by a mechanism other than controlling cell proliferation.

Another function of Rho-family GTPases is to control cytoskeletal organization in response to extracellular factors¹². Cytoskeletal proteins are known effectors for events essential for cell motility²⁵. Therefore, RhoC may control metastasis by regulating cell motility. Metastatic A375M cells were more migratory (Fig. 3a) and more invasive (Fig. 3b) than the poorly metastatic A375P cells. Furthermore, RhoC could enhance the migratory and invasive capacity of the A375P cells, whereas dnRho inhibited motility and invasion of the A375M cells, indicating that RhoC may regulate metastasis by controlling cytoskeletal events essential for motility. We also observed that RhoC could induce in A375P cells an elongated morphology similar to that of A375M cells, while dnRho expression suppressed this morphology (Fig. 4). Another morphological difference noted in the A375M cells, the serum-induced formation of filopodia, also correlated with the metastatic capacity of these cells. However, filopodial protrusions (which are regulated by the Rho subfamily member Cdc42 (ref. 12) were not altered by expression of RhoC or dnRho (data not shown), indicating that regulation of these actin-based structures may occur upstream (or independently) of Rho.

We have identified RhoC as essential for tumour metastasis. Compared with RhoA, the canonical family member, little is

Table 2 Pulmonary metastases

Cell line	No. of metastases	No. of mice
A375P	0,0,0,0,1,5,10	8
A375P-RhoC	56,70,>100,>100	4
A375M	all >100	8
A375M-dnRho	13,24,29,32	4

Numbers of pulmonary metastases identified on the surface of the lungs of mice injected with A375P, A375P-RhoC, A375M or A375M-dnRho cells.

known about RhoC. RhoA and RhoC are highly homologous, with only six non-conservative amino-acid substitutions, all in the carboxy-terminal end of the molecules. It might be thought that RhoA should be able to enhance tumour metastasis. However, RhoA is expressed at equivalent levels in both the poorly and highly metastatic tumours (see Supplementary Information), indicating that the level of RhoA expression in the A375 cells is not sufficient for metastasis. Furthermore, when expressed at equivalent levels, RhoC was a better motogen than was RhoA (Fig. 3). These results indicate that there may be a functional difference between Rho subfamily members that requires further investigation. Finally, the observation that expression of a single gene is sufficient to induce metastasis is perhaps surprising, given that metastasis is such a complex process. We suspect, however, that many cells within the heterogeneous tumorigenic A375P population may be genetically primed for metastasis so that introduction of a single gene (such as RhoC) which affects a process essential for metastasis is sufficient for metastasis. We are currently examining whether RhoC is capable of inducing metastasis in other tumorigenic cells. □

Methods

Cell lines

The A375 (ATCC#CRL-1619) and B16 (ATCC#CRL-6322) cell lines were maintained as described⁷. Cells were harvested by trypsinization, washed in PBS and diluted to 2.5×10^6 cells per ml for A375 cells and 2.5×10^5 cells per ml for B16 cells.

Experimental metastasis assay

A375 cells were injected either intravenously (0.2 ml) into the lateral tail vein or subcutaneously (0.1 ml) into the dorsal flank of nude mice, and B16 cells were injected into syngeneic C57BL/6 mice. Three (for B16) to eight (for A375) weeks after injection the mice were killed; the lungs were removed and washed and the pulmonary metastases on the lung surface were counted under a dissecting microscope. Metastatic nodules were removed aseptically, minced and grown *in vitro*, or snap-frozen in liquid nitrogen to purify RNA.

Tumours and tumour-derived cell lines

A375M1, M2 and SM lines were selected using the experimental metastasis assay for their enhanced ability to form experimental pulmonary metastases⁶. Line M1 was derived from metastases isolated from mice injected intravenously with the A375P cells, line M2 from mice injected with A375M1 cells, and line SM was a gift from I. Fidler and was derived by an identical selection procedure⁷. B16 lines were derived in an identical manner, with F1 cells derived from B16F0 cells, F2 from B16F1 cells and F3 from B16F2 cells. The A375M cell line is a pool of cells from A375M1, M2 and SM cells. A375P and A375M cells used in retroviral gene transfer studies were transfected with a plasmid containing the ecotropic receptor (a gift from H. Lodish) and selected for neomycin-resistance. Mock-infected or uninfected cells were used as negative controls in the metastasis assays.

Array hybridization

Total RNA was prepared with a Qiagen RNeasy mini-kit according to the manufacturer's instructions. We prepared cRNA for hybridization essentially as described²⁶. Oligo-nucleotide arrays (GeneChip, Affymetrix) composed of 7,070 human (HUM 6.8K) or 6,347 mouse (MUR 6K) genes and ESTs were used for hybridization according to the manufacturer's instructions. Arrays were scanned using an Affymetrix confocal scanner and analysed using GeneChip 3.0 software (Affymetrix). Intensity values were scaled so that the overall fluorescence intensity of each chip of the same type was equivalent. For a gene to be selected as induced, it has to be expressed in all three metastatic samples at least 2.5 times higher than in the poorly metastatic sample, with experiments done in duplicate. Where expression in the poorly metastatic sample was below baseline (set at 20, the point below which changes in expression could be determined with high confidence), it was determined to be absent and was set to 20.

Retroviral gene transfer

An EcoRI fragment of pCR-BluntII-RhoC containing the entire coding region of human RhoC (see Supplementary Information) was inserted into the EcoRI site of the retroviral bicistronic expression vector pMX-IRES-GFP (pMIG; ref. 27) containing enhanced green fluorescent protein (GFP) as an expression marker. An EcoRI fragment of pEXV-RhoA or pEXV-N19RhoA (a dominant-negative Rho mutant, dnRho) was inserted into the EcoRI site of pMIG. We transfected pMIG-RhoC, pMIG-RhoA and pMIG-dnRho into 293T cell-derived retroviral producer lines (Phoenix cells) as described (see www.stanford.edu/group/nolan/) and A375P or M cells were infected. Cells were sorted by FACStar (Becton-Dickinson) according to their GFP levels. RNAse protection assays showed that RhoC was expressed at 5- to 50-fold higher levels than in the A375M cells, whereas RhoA and RhoC were expressed at similar levels in the A375P-RhoC and A375P-RhoA cells (as determined by GFP expression).

Chemotaxis and invasion assays

Cell migration and invasion assays were performed using 8.0- μ m pore size Transwell inserts (Costar Corporation) or Biocoat Matrigel invasion chambers (Becton-Dickinson), respectively²⁸. Each data point represents the average of 3 or 4 individual experiments, done in duplicate, and error bars represent the standard error of the mean.

Immunofluorescence

Adherent cells were fixed, permeabilized and stained as described²⁹.

Received 31 January; accepted 15 May 2000.

1. Van Noorden, C. J. F., Meade-Tollin, L. C. & Bosman, F. T. Metastasis. *Am. Sci.* **86**, 130–141 (1998).
2. Fidler, I. J. & Radinsky, R. Search for genes that suppress cancer metastasis. *J. Natl Cancer Inst.* **88**, 1700–1703 (1996).
3. Roberts, D. D. Regulation of tumor growth and metastasis by thrombospondin-1. *FASEB J.* **10**, 1183–1191 (1996).
4. Bao, L. *et al.* Thymosin β 15: a novel regulator of tumor cell motility upregulated in metastatic prostate cancer. *Nature Med.* **2**, 1322–1328 (1996).
5. Habets, G. G. M. *et al.* Identification of an invasion-inducing gene, Tiam-1, that encodes a protein with homology to GDP-GTP exchangers for rho-like proteins. *Cell* **77**, 537–549 (1994).
6. Fidler, I. J. Selection of successive tumour lines for metastasis. *Nature* **242**, 148–149 (1973).
7. Kozlowski, J. M., Hart, I. R., Fidler, I. J. & Hanna, N. A human melanoma cell line heterogeneous with respect to metastatic capacity in athymic nude mice. *J. Natl Cancer Inst.* **72**, 913–917 (1984).
8. Welch, D. R. *et al.* Microcell-mediated transfer of chromosome 6 into metastatic human C8161 melanoma cells suppresses metastasis but does not inhibit tumorigenicity. *Oncogene* **9**, 255–262 (1994).
9. Zhang, L. *et al.* Gene expression profiles in normal and cancer cells. *Science* **276**, 1268–1272 (1997).
10. Maniotis, A. J. *et al.* Vascular channel formation by human melanoma cells *in vivo* and *in vitro*: vasculogenic mimicry. *Am. J. Pathol.* **155**, 739–752 (1999).
11. Humphries, M. J., Olden, K. & Yamada, K. M. A synthetic peptide from fibronectin inhibits experimental metastasis of murine melanoma cells. *Science* **233**, 467–469 (1986).
12. Van Aelst, L. & D'Souza-Schorey, C. Rho GTPases and signaling networks. *Genes Dev.* **11**, 2295–2322 (1997).
13. Suwa, H. *et al.* Overexpression of the rhoC gene correlates with progression of ductal adenocarcinoma of the pancreas. *Br. J. Cancer* **77**, 147–152 (1998).
14. Hall, A. K. Differential expression of thymosin genes in human tumors and in the developing human kidney. *Int. J. Cancer* **48**, 672–677 (1991).
15. Weterman, M. A. J., van Muijen, G. N. P., Ruiter, D. J. & Bloemers, H. P. J. Thymosin β -10 expression in melanoma cell lines and melanocytic lesions: a new progression marker for human cutaneous melanoma. *Int. J. Cancer* **53**, 278–284 (1993).
16. Chen, L., O'Bryan, J. P., Smith, H. S. & Liu, E. Overexpression of matrix Gla protein mRNA in malignant human breast cells: isolation by differential cDNA hybridization. *Oncogene* **5**, 1391–1395 (1990).
17. Schonherr, E. *et al.* Interaction of biglycan with type 1 collagen. *J. Biol. Chem.* **270**, 2776–2783 (1995).
18. Svensson, L. *et al.* Fibromodulin-null mice have abnormal collagen fibrils, tissue organization, and altered luminal deposition in tendon. *J. Biol. Chem.* **274**, 9636–9647 (1999).
19. Ruoslahti, E. Fibronectin and its integrin receptors in cancer. *Adv. Cancer Res.* **76**, 1–20 (1999).
20. Jeffers, M., Rong, S. & Vande Woude, G. F. Enhanced tumorigenicity and invasion-metastasis by hepatocyte growth factor/scatter factor-Met signalling human cells concomitant with induction of the urokinase proteolysis network. *Mol. Cell. Biol.* **16**, 1115–1125 (1996).
21. Chambers, A. F. & Matrisian, L. M. Changing views of the role of matrix metalloproteinases in metastasis. *J. Natl Cancer Inst.* **89**, 1260–1270 (1997).
22. Albelda, S. M. *et al.* Integrin distribution in malignant melanoma: association of the β 3 subunit with tumor progression. *Cancer Res.* **50**, 6757–6764 (1990).
23. Quilliam, L. A., Khosravi-Far, R., Huff, S. Y. & Der, C. J. Guanine nucleotide exchange factors: activators of the Ras superfamily of proteins. *BioEssays* **17**, 395–404 (1995).
24. Feig, L. A. & Cooper, G. M. Inhibition of NIH 3T3 cell proliferation by a mutant ras protein with preferential affinity for GDP. *Mol. Cell. Biol.* **8**, 3235–3243 (1988).
25. Lauffenburger, D. A. & Horowitz, A. F. Cell migration: a physically integrated molecular process. *Cell* **84**, 359–369 (1996).
26. Fambrough, D., McClure, K., Kazlauskas, A. & Lander, E. S. Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. *Cell* **97**, 727–741 (1999).
27. Liu, X. *et al.* Transforming growth factor β -induced phosphorylation of Smad3 is required for growth inhibition and transcriptional induction in epithelial cells. *Proc. Natl Acad. Sci. USA* **94**, 10669–10674 (1997).
28. Yebra, M. *et al.* Requirement of receptor-bound urokinase-type plasminogen activator for integrin α v β 5-directed cell migration. *J. Biol. Chem.* **271**, 29393–29399 (1996).
29. Clark, E. A., King, W. G., Brugge, J. S., Symons, M. & Hynes, R. O. Integrin-mediated signals regulated by members of the rho family of GTPases. *J. Cell Biol.* **142**, 573–586 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank C. Huard, C. Whittaker, S. Robinson, J. Lively, D. Hirsch, D. Crowley and P. Tamayo for technical assistance and advice, and H. Lodish, G. Nolan and J. Fidler, for reagents. This work was supported in part by grants from the National Cancer Institute (to R.O.H.), Affymetrix, Inc., Bristol-Myers Squibb and Millenium Pharmaceuticals (to E.S.L.), and a Merck/MIT postdoctoral fellowship (to E.A.C.). R.O.H. is an investigator and E.A.C. was an associate of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to R.O.H. (e-mail: rohynes@mit.edu).

Molecular classification of cutaneous malignant melanoma by gene expression profiling

M. Bittner^{††}, P. Meltzer^{††}, Y. Chen^{*}, Y. Jiang^{*}, E. Seftor[‡], M. Hendrix[‡], M. Radmacher[§], R. Simon[§], Z. Yakhini^{||}, A. Ben-Dor^{||}, N. Sampas^{||}, E. Dougherty[#], E. Wang^{*}, F. Marincola^{*}, C. Gooden^{*}, J. Lueders^{*}, A. Glatfelter^{*}, P. Pollock^{**}, J. Carpten^{*}, E. Gillanders^{*}, D. Leja^{*}, K. Dietrich^{*}, C. Beaudry^{††}, M. Berens^{††}, D. Alberts^{‡‡}, V. Sondak^{§§}, N. Hayward^{**} & J. Trent^{*}

^{*} Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892, USA

[‡] Department of Anatomy and Cell Biology, University of Iowa Cancer Center, Iowa City, Iowa 52242-1109, USA

[§] National Cancer Institute, DCTDC, NIH, Bethesda, Maryland 20852, USA

^{||} Chemical and Biological Systems Department, Agilent Laboratories, 3500 Deer Creek Road, Palo Alto, California 94304, USA

[#] Computer Science and Engineering Department, University of Washington, Seattle, Washington 98105, USA

[#] Department of Electrical Engineering, Texas A & M University, College Station, Texas 77843, USA

^{*} National Cancer Institute, Surgery Branch, NIH, Bethesda, Maryland 20850, USA

^{**} Queensland Institute of Medical Research, Herston, Queensland 4029, Australia

^{††} Neuro-Oncology Laboratory, Barrow Neurological Institute, Phoenix, Arizona 85013-4496, USA

^{‡‡} Arizona Cancer Center, University of Arizona, Tucson, Arizona 85724, USA

^{§§} Department of Surgery, University of Michigan, Ann Arbor, Michigan 48109, USA

[†] These authors contributed equally to this work.

The most common human cancers are malignant neoplasms of the skin^{1,2}. Incidence of cutaneous melanoma is rising especially steeply, with minimal progress in non-surgical treatment of advanced disease^{3,4}. Despite significant effort to identify independent predictors of melanoma outcome, no accepted histopathological, molecular or immunohistochemical marker defines subsets of this neoplasm^{2,3}. Accordingly, though melanoma is thought to present with different 'taxonomic' forms, these are considered part of a continuous spectrum rather than discrete entities². Here we report the discovery of a subset of melanomas identified by mathematical analysis of gene expression in a series of samples. Remarkably, many genes underlying the classification of this subset are differentially regulated in invasive melanomas that form primitive tubular networks *in vitro*, a feature of some highly aggressive metastatic melanomas⁵. Global transcript analysis can identify unrecognized subtypes of cutaneous melanoma and predict experimentally verifiable phenotypic characteristics that may be of importance to disease progression.

We, and others⁶⁻¹⁰, have proposed that a discrete and previously unrecognizable cancer taxonomy can be identified by viewing the systematized data from gene expression experiments. However, for melanoma, inherent or technically induced variation could obscure such a classification as its appearance is very similar between patient samples and, in contrast to haematologic cancers^{9,10}, it has few known recurring genetic changes. To explore this question, we gathered expression profiles for 38 samples, including 31 melanomas and 7 controls (Table 1). We isolated total messenger RNA directly from melanoma biopsies or tumour cell cultures, prepared fluorescent complementary DNA from the message and hybridized them to a microarray containing probes for 8,150 cDNAs (representing 6,971 unique genes), obtaining quantitative and compara-

tive measurements for each gene.

We compared the tumour cell mRNA with a single reference probe, providing normalized measures of the expression of each gene in each sample relative to the standard. Analysis of the normalized expression across all genes between samples provided a measure of the overall difference in expression pattern between samples. Similarly, the orthogonal analysis of linear covariance between pairs of genes across all samples provided a measure of the similarity of behaviour of the genes studied.

Figure 1 shows the integration of several analytical methods to visualize the overall expression pattern relationships between cutaneous melanoma tumour samples. Using a matrix of Pearson correlation coefficients from the complete pair-wise comparison of all experiments¹¹, the 31 melanoma experiments are displayed as a hierarchical clustering dendrogram^{7,12} and as a three-dimensional multidimensional scaling (MDS) plot^{7,13}. The MDS plot displays the position of each tumour sample in three-dimensional Euclidean space, with the distance between experimental samples reflecting their approximate degree of correlation^{7,13}. The analysis included all genes meeting a minimum level of expression in each hybridization. We also employed a non-hierarchical clustering algorithm (termed cluster affinity search technique; CAST)¹⁴ to define experimental clusters. The resulting hierarchical dendrogram of the 31 melanoma samples (Fig. 1a) demonstrates that 19 samples are tightly clustered at the bottom of the dendrogram in the area of highest similarity. Likewise, the non-hierarchical CAST algorithm identified the identical major cluster 19 melanomas. This cluster is also a compact, readily separable grouping based on its overall similarity of expression pattern viewed by MDS (Fig. 1b). A three-dimensional animation of the MDS plot is presented in the Supplementary Information.

There is no single established method to estimate the significance of an observed degree of relationship obtained by cluster prediction techniques^{9,11}. Accordingly, we used two independent approaches to test the validity of our cluster prediction of the 19-element cluster. The first approach (Fig. 1c) examines the power of individual genes to discriminate the major cluster of 19 from the remaining samples by examining the frequency of strong classifier genes compared to the expected frequency of such genes if expression is randomly variable, and to the frequency of strong classifiers in random partitions of the same samples into new groupings of 19 and 12. The non-randomness of the cluster results is evident. Specifically, many genes have expression patterns that differ strongly between the initial sample clusters and thus serve as good classifiers (Fig. 1c, red triangles). However, expression patterns are not readily found which classify the samples when they are grouped into random partitions of the same size (Fig. 1c, blue lines). Accordingly, in randomly formed clusters, expression behaviour is essentially indistinguishable from truly random behaviour of genes relative to these clusters (Fig. 1c, compare blue lines with open circles).

The second approach we used to test the validity of the cluster predictions is based on evaluating cluster membership after introducing random perturbations to the data set. For each sample, the log-ratio of each gene was perturbed by the introduction of random gaussian noise with the mean equal to 0 and the standard deviation equal to 0.15 (an estimate of variation derived by computing the median standard deviation of the log-ratios for single genes across all 31 samples). Hierarchical clustering was then performed on the perturbed data set and a comparison made between the original tree (Fig. 1a) and the perturbed tree. Comparisons involved cutting the original and perturbed trees into *k* clusters followed by computing the proportion of paired samples clustering together in the original tree that did not cluster together in the perturbed tree (we refer to this measure as a weighted proportion of discrepant pairs because it gives more weight to larger clusters). The comparison was repeated over multiple perturbed data sets for each possible cut in the

original tree ($k = 2, 3, \dots, 30$). For a given k , the weighted proportion of discrepant pairs was then averaged over the perturbed data sets resulting in the identification of weighted average discrepant pairs (WADP_k; see Supplementary Information).

Clusters that result from cutting the original tree into 9 or fewer groups are very reproducible (Fig. 1d). It is noteworthy that the rise in WADP_k almost exactly coincides with the division of the major 19-element cluster into smaller sub-clusters. These results strongly support the view that the major cluster of melanoma samples identified in this study represents a bona fide and highly reproducible grouping.

We then performed statistical tests to determine whether any clinical or tumour cell characteristics were specifically associated with the clustered group. Tests for associations between the major cluster of 19 samples and the remaining 12 melanoma samples were performed for several *in vivo* variables, including sex, age, biopsy site, Breslow thickness, Clark's level and survival. There was no statistically significant association between the cluster group and any clinical variable. There were also no significant associations with the *in vitro* variables, including p16 or β -catenin mutation status, *in vitro* pigmentation and cell passage number (see Supplementary Information).

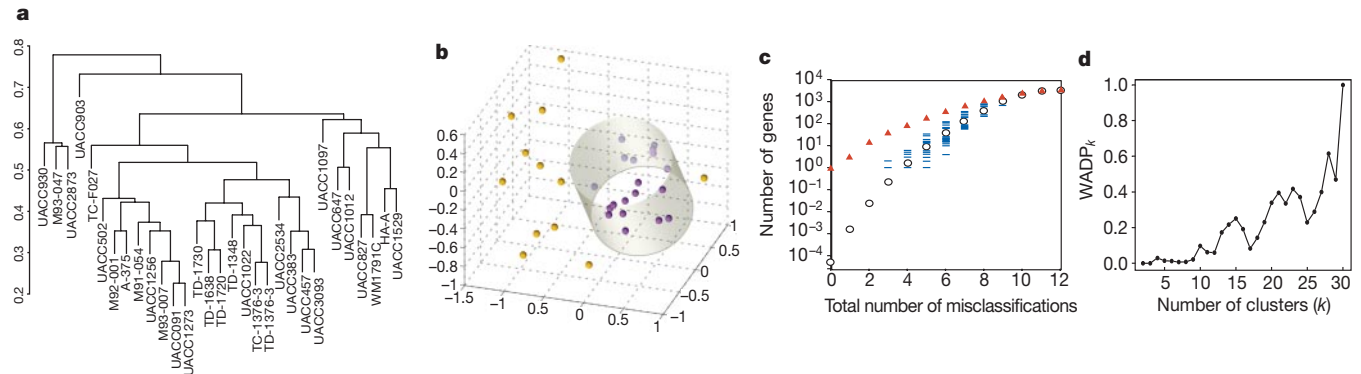


Figure 1 Clustering of gene expression data. **a**, Hierarchical clustering dendrogram with the cluster of 19 melanomas at the centre. **b**, MDS three-dimensional plot of all 31 cutaneous melanoma samples showing major cluster of 19 samples (blue, within cylinder), and remaining 12 samples (gold). See Supplementary Information to view a three-dimensional animation. **c**, A plot of the observed and expected number of genes producing a given number of classification errors for a partition of the 31 melanomas into

two groups of 12 and 19. Red triangles, observed clusters; filled bars, randomly produced clusters; open circles, predicted results for randomly variable gene expression. **d**, Introduction of random gaussian noise followed by cuts from the top of the original tree (resulting in k clusters), to determine discrepant pairs after perturbation (see Supplementary Information).

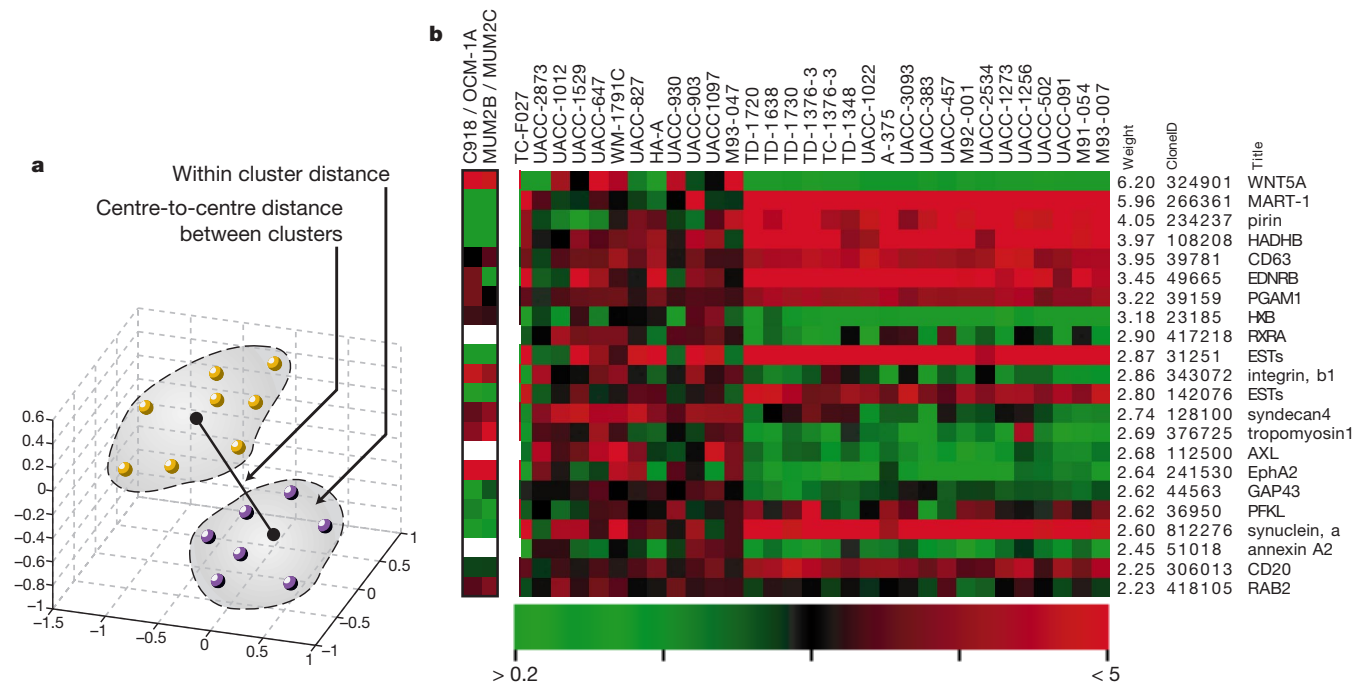


Figure 2 Identifying genes which discriminate melanoma clusters. **a**, MDS analysis ranking genes according to their impact on minimizing cluster volume and maximizing centre-to-centre inter-cluster distance. **b**, Top 22 genes obtained by these criteria listed in order of decreasing weight (for a full list see Supplementary Information). Right, data from cutaneous melanomas identified on the horizontal axis and sorted by cluster

(described in ref. 5). Left, data from uveal melanomas expressed as the ratio of highly invasive to less invasive. Red, high ratios; green, low ratios (intensity of saturation scaled according to the ratio). The three genes not scored in the uveal samples were not included in the print design of the cutaneous samples.

We included two pairs of specimens derived from the same patient in this sample set. These are M92-001 and M93-007 (two different samples from the same individual, surgically removed one year apart), and TD-1376-3 and TC-1376-3 (the biopsy sample and a cell culture of the same tumour carried three passages *in vitro*). Although there was no significant association between cell passage number and cluster group ($P = 0.857$, see Supplementary Information), the TD-1376-3/TC-1376-3 pair were included to serve as another control for the effects of cell culture. Remarkably, of the 465 pairwise comparisons among the melanoma samples, the pairs TD-1376-3/TC-1376-3 and M92-001/M93-007 are the second and third most highly correlated pairs of samples, with nearly identical correlation coefficients (Fig. 1b).

On the basis of the linear correlation of global gene expression in Fig. 1, Figs 2 and 3 illustrate the approach we have used to guide 'gene cluster' interpretation empirically. Fig. 2a depicts our statistical method for extracting a 'weighted list' of individual genes whose variance of change across all experiments correctly defines the boundary of a given sample cluster (for details see Supplemen-

tary Information). Fig. 2b displays the list of genes with the most power to define the major melanoma cluster of 19 samples (Fig. 1a and b) in rank order along the vertical axis. The samples are ordered along the horizontal axis by cluster inclusion, and data are presented graphically as coloured images with the colour saturation directly proportional to the magnitude of the measured gene expression ratio (brightest red, highest R/G ratio; black squares, R/G ratio = 1; brightest greens, lowest R/G ratio). The complete list of genes discriminating the major cluster is in the Supplementary Information.

The weighted gene list can also be used to guide analysis of the larger gene expression data set. Figure 3a displays all data from the cutaneous melanoma samples in this study as a coloured image with genes ordered along the vertical axis by similarity of expression pattern (after ref. 12). However, rather than basing analysis of this large (> 300,000 elements) data set entirely on visual selection, we used genes from the weighted list to index gene cluster selection. Figure 3b–e illustrates this approach using four genes from the 'weighted list' in Fig. 2b (MART-1, CD63, tropomyosin and

Table 1 Summary of melanoma cases by cluster designation

Case no.	Sex/Age	Biopsy site	Passage no. (Biopsy)	p16 mutation status*	Invasive ability†	Vasulogenic mimicry‡	Gel contraction§	Cell motility	Scratch wound (%)¶
Melanoma primary cluster									
UACC-502	M/69	Cervical node	3	Deleted	2.8 ± 0.1%	–	ND	ND	37
M92-001	F/43	Ankle	2	Deleted	3.0 ± 0.5%	–	ND	76.80 ± 2.96	22
A-375	F/54	Skin	ND	Mutation	2.8 ± 0.2%	–	ND	67.80 ± 4.40	26
M91-054#	M/45	Axill. lymph node	3	WT	#	#	#	ND	30
UACC-1256	F/67	Thigh femoral node	9	Deleted	ND	ND	ND	ND	ND
M93-007	F/43	Ankle	3	Deleted	2.6 ± 0.1%	–	–	ND	12
UACC-091	M/52	Unk	7	Deleted	2.1 ± 0.2%	–	–	ND	11
UACC-1273	M/50	Axill. lymph node	16	Mutation	2.5 ± 0.3%	–	–	ND	13
TD-1730	M/55	Thyroid lobe	Biopsy	ND	ND	ND	ND	ND	ND
TD-1638	M/49	Paraspinous	Biopsy	ND	ND	ND	ND	ND	ND
TD-1720	M/29	Shoulder	Biopsy	ND	ND	ND	ND	ND	ND
TD-1348	M/44	Axill. lymph node	Biopsy	ND	ND	ND	ND	ND	ND
UACC-1022	F/53	Chest wall	13	WT	2.9 ± 0.1%	–	–	ND	63
TC-1376*	M/30	Distal ileum	3	ND	ND	ND	ND	ND	21
TD-1376*	M/30	Distal ileum	Biopsy	ND	ND	ND	ND	ND	ND
UACC-2534	M/68	Abdomen	7	Deleted	3.2 ± 0.02%	–	ND	ND	7
UACC-383	M/69	Thigh femoral node	29	Deleted	2.3 ± 0.2%	–	ND	70.40 ± 5.27	35
UACC-457	F/Unk	Unk	19	WT	3.1 ± 0.2%	–	ND	12.80 ± 0.05	ND
UACC-3093	M/75	Axill. lymph node	4	WT	ND	ND	ND	40.30 ± 2.00	24
Melanoma non-clustered									
UACC-930	F/35	Sm. bowel	4	WT	4.8 ± 0.3%	±	–	ND	50
M93-047	F/75	Axill. lymph node	3	Mutation	10.7 ± 0.03%	+	+	ND	75
UACC-2973	M/37	Axill. lymph node	5	ND	ND	ND	ND	ND	48
UACC-903	M/25	Back	14	Deleted	3.8 ± 0.3%	+	–	ND	91
TC-FO27	M/30	Rt. chest wall	6	ND	ND	ND	ND	ND	91
UACC-1097	M/56	Rectus muscle	6	Mutation	ND	ND	ND	ND	34
UACC-647**	M/32	Axill. node	14	WT	3.8 ± 0.1%	+	±	ND	55
UACC-1012	M/54	Neck	3	ND	4.9 ± 0.1%	ND	ND	122.00 ± 11.30	54
UACC-827	F/32	Rt. breast	16	Mutation	ND	ND	ND	ND	32
WM1791C	Unk	Unk	52	ND	4.6 ± 0.3%	+	ND	141.00 ± 11.40	71
HA-A	F/Unk	Unk	19	ND	3.9 ± 0.5%	±	ND	211.00 ± 12.40	62
UACC-1529	M/48	Axill. lymph node	13	Mutation	4.2 ± 0.5%	+	–	ND	ND
Uveal melanoma samples									
OCM-1A	Unk	Primary	25	ND	2.2 ± 0.1%	–	–	ND	ND
C918	F/60	Primary	15	ND	12.9 ± 0.3%	+	+	ND	ND
MUM-2C	M	Liver metastasis	8	ND	2.0 ± 0.1%	–	–	ND	ND
MUM-2B	M	Liver metastasis	8	ND	13.3 ± 0.6%	+	+	ND	ND
Control samples									
Nil. C (fibroblast); UACC-3149 (ovarian adenocarcinoma); MCF-10A (breast epithelium); CRL-1634 (fibroblast); SRS-3 (cell culture variant); SRS-5 (cell culture variant); RMS-13 (rhabdomyosarcoma)									

* Mutation status of indicated samples for p16 obtained by sequencing. Deleted, homozygous. Supplementary Information includes the specific mutations in p16 for each sample tested. Samples were also sequenced for β -catenin. No example of β -catenin mutation was observed.

† Ability to invade a defined basement matrix. $P = 0.0055$; t -test for two populations.

‡ Tube forming ability at 5 days in a three-dimensional matrigel matrix.

§ Ability to contract floating collagen I gels at 5 days as compared to HT-1080 fibrosarcoma cells⁵.

|| Migration rates expressed in μ m per day. Mean from eight experiments $\pm$ s.d. ($P = 0.0063$; t -test for two populations). Rates below 100 μ m per day completely segregates in the melanoma primary cluster.

¶ Ability to close *in vitro* scratch wound at 24 h. Photographs of the wound were measured and percentage wound closure determined²³ ($P < 0.00002$, t -test for two populations).

M91-054 was the only sample that demonstrated a mixed phenotype in culture with both an epithelioid population and a more fibroblastic population. Vasculogenic mimicry and gel contraction were only observed in the epithelioid population. Scratch assay resulted in 30% closure after 24 h for both populations.

* TC-1376 mRNA was isolated after short term (3 passage) culture of the biopsy sample from the patient TD-1376 allowing the effects of short term culture on the expression profile to be observed.

** UACC-647 cells form extensive cord-like networks by 5 days.

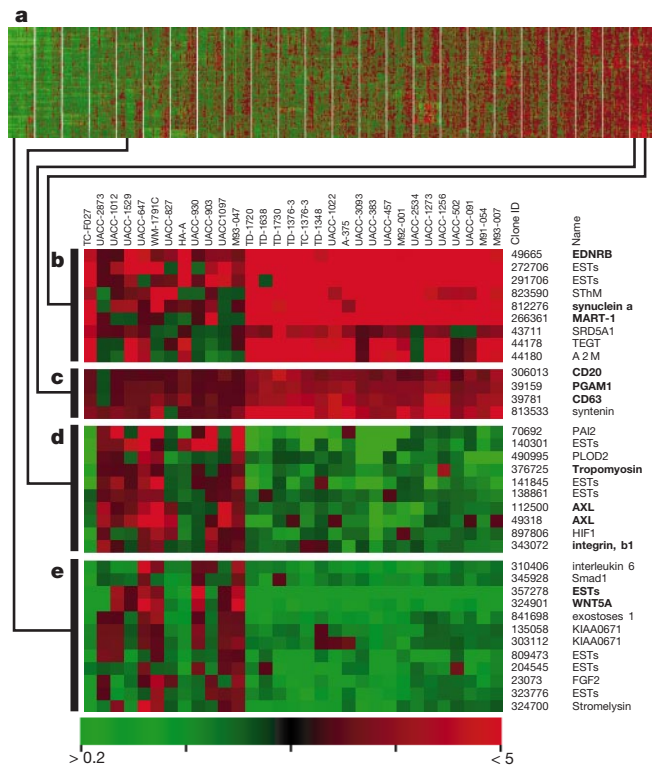


Figure 3 Guiding gene cluster selection. **a**, Two-dimensional cluster analysis of cutaneous melanoma samples (horizontal axis) and genes (vertical axis, presented in segments). **b–e**, Data from **a** queried at regions corresponding to four top discriminators of the major cluster: MART-1 (**b**), CD63 (**c**), tropomyosin (**d**) and WNT5A (**e**). Note that these clusters include other genes from the discriminator list (bold). The major cluster of 19 samples is visually apparent on the left of this display. The full list of gene names and corresponding calculated ratio information is provided in the Supplementary Information.

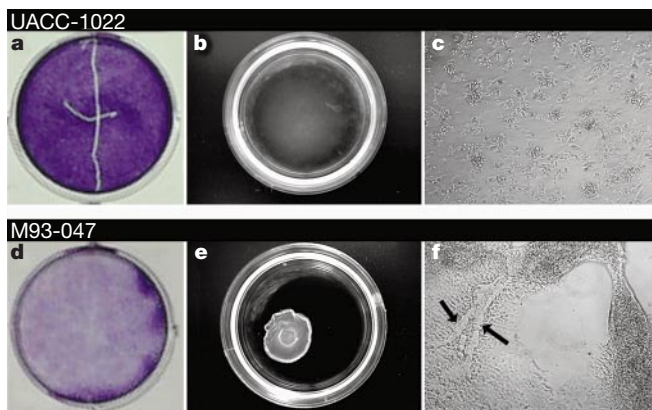


Figure 4 Variation in biological properties of melanoma clusters. **a–c**, A representative member of the major melanoma cluster (UACC-1022). **d–f**, A sample falling outside of the major cluster (M93-047). The two groups differ in the ability to migrate into a scratch wound (**a**, **d**), contract collagen gels (**b**, **e**) and form tubular networks (**c**, **f**). Results of these and additional cell mobility/invasion assays are included in Table 1. Tubular network formation (vasculogenic mimicry⁵, **f**) and collagen gel contraction (related to the patterning of vascular channels, **e**) were observed only outside the major cluster (Table 1).

WNT5A), to interrogate the entire gene expression data set represented in Fig. 3a.

Finally, in parallel to our microarray analysis of cutaneous melanoma, we studied a series of uveal melanoma specimens characterized for properties related to metastasis, including invasive ability and vasculogenic mimicry *in vitro*⁵. These samples were hybridized pairwise, directly comparing highly invasive cells to their less invasive counterparts. We examined the pattern of gene expression in these phenotypically characterized cells with respect to the weighted discriminator list (Fig. 2b) that defines the major cluster of 19 cutaneous melanomas. Strikingly, genes expressed in common in the highly invasive uveal melanoma cells (Fig. 2b, inset) were strongly anti-correlated with the same gene from the major cluster of cutaneous melanoma samples (Fig. 2b). This observation, coupled with the known biological function of genes within the weighted list, indicated that specimens assigned within the major cutaneous melanoma cluster (Fig. 1a, b) would have reduced motility and reduced invasive ability as they have down-regulation of genes related to cell spreading or migration, including formation of focal adhesions^{15,16}. Specific genes with reduced expression in the major cluster included integrin $\beta 1$ (refs 17, 18), integrin $\beta 3$ (ref. 19), integrin $\alpha 1$ (ref. 18), syndecan 4 (ref. 20) and vinculin²¹ (Figs 2 and 3; see Supplementary Information). In samples outside the major cluster increased expression of fibronectin is particularly interesting. With other reports^{22,23}, this observation indicates that these cells are induced to secrete this pro-migratory molecule, consistent with an important role for focal contacts in modulating melanoma cell motility.

We then directly tested the prediction from the array results that cell spreading and migration could be discordant between melanoma cluster groups. Cutaneous melanomas (assigned either in or out of the major cluster) were characterized using a series of cellular assays applied to test cell motility and invasiveness (Table 1, Fig. 4). Figure 4 illustrates the discordance of cutaneous melanoma samples within the major cluster and those outside this group. As predicted from the analysis of their gene expression patterns, melanomas within the major cluster had reduced motility ($P = 0.0063$), invasive ability ($P = 0.0055$) and vasculogenic mimicry in comparison with melanomas outside the major cluster (Table 1).

The patient population in this study had a uniformly poor prognosis, and neither typical clinical factors (for example, age, sex, biopsy site) nor *in vitro* characteristics (for example, passage number) provide strong correlation with clinical outcome, or expression information (see Supplementary Information). In contrast, molecular classification of these tumours on the basis of gene expression (Fig. 1, Table 1) could identify a previously undetected subtype of this cancer. The analyses described here were not designed to address the relationship of gene expression profile and clinical outcome in melanoma patients, and thus the clinical relevance of our observed subgrouping awaits further analysis. However, survival information was available on 15 patients, and the results, though not statistically significant, are of interest. Three deaths occurred out of 10 patients in the tight cluster of 19 while 4 deaths occurred out of 5 patients in the remaining group (log-rank P -value = 0.135). Our results indicate melanoma will provide a unique opportunity to study a homogeneous group of patients to determine if gene expression patterns predict prognosis or therapeutic response in settings where we cannot currently determine who is most at risk for rapid disease progression and death.

Finally, classification of melanoma on the basis of gene expression patterns is possible, despite the prevailing view that the 'taxonomy' of this disease falls in a continuous spectrum lacking discernable entities². Our data show that melanoma is a useful model to identify genes critical for aspects of the metastatic process, including tumour

cell motility and the ability to form primitive tubular networks that may contribute to tumour perfusion. The extent to which melanoma samples can be clinically subdivided by expression patterns remains to be elucidated. However, our identification of genes 'weighted' for their ability to discriminate a subset of melanomas should provide a sound molecular basis for the dissection of other clinically relevant subsets of this tumour. □

Methods

Samples

Cultured cells were collected and mRNA isolated as described (ref. 25; www.nhgri.nih.gov/DIR/microarray). Samples underwent a series of controls for quality of mRNA, labelling and hybridization, as well as sample integrity (including genotyping DNA from all samples with five dinucleotide markers from four different chromosomes to insure individuality). The entire coding sequence of the p16 gene and exon 3 of the β -catenin genes was sequenced to assess the mutation status of all available samples (see Supplementary Information). The biopsy tumour specimens used in this study were obtained with Institutional Review Board approval and clinical information is provided in the Supplementary Information. Biopsies were debried, dissected into small pieces and frozen in liquid nitrogen. Frozen specimens were immediately placed into TRIzol Reagent (Gibco BRL), homogenized and mRNA isolated as described (ref. 25; www.nhgri.nih.gov/DIR/microarray).

Microarrays

The 8,150 human cDNAs used in this study were obtained under a Cooperative Research and Development Agreement with Research Genetics and 6,912 were verified by sequence. This set of cDNAs is part of a larger collection (refs 7, 24; www.nhgri.nih.gov/DIR/microarray). On the basis of the Unigene build of 9 March 2000 (<http://www.ncbi.nlm.nih.gov/UniGene/build.html>), the 8,150 cDNAs represent 6,971 unique genes in this melanoma array. All clones were confirmed by resequencing if necessary. Microarrays were hybridized, scanned and image analysis performed as described (refs 7, 25; www.nhgri.nih.gov/DIR/microarray).

Statistical methods

Detailed information on all statistical methods is in the Supplementary Information. Agglomerative hierarchical clustering of the 31 melanomas on the basis of their gene expression profiles was performed as described^{7,11}, to investigate relationships between tumour samples. Average linkage was used, as well as a dissimilarity measure of one minus the Pearson correlation coefficient of log ratios. The cutoff employed to obtain the observed partitioning was 0.54. The MDS was performed using an implementation of MDS in the MATLAB package. A non-hierarchical clustering algorithm¹⁴ was used to define experimental clusters. This approach takes a graph theoretic approach, and makes no assumptions on the similarity function or the number of clusters sought.

To generate the weighted gene list, cluster compaction and separation were evaluated. For a given clustering result, $n_1 = 19$ and $n_2 = 12$, the discriminative weight of each gene $w = d_B/(k_1 d_{w1} + k_2 d_{w2} + \alpha)$; where d_B is the centre-to-centre distance (between cluster Euclidean distance), d_{wi} is the average Euclidean distance among all sample pairs within cluster i , $k_i = t_i/(t_1 + t_2)$ for a total of t_i sample pairs in cluster i , and α is a small constant (0.1 in our study) to prevent the zero denominator case (Fig. 2a). Genes may then be ranked on the basis of w .

In vitro biological assays

Floating collagen lattices were prepared and used to test selected cell lines for their ability to deform the gels as described (ref. 5; Table 1 legend). Samples were also tested for their ability to migrate into an *in vitro* scratch wound as described²⁶. Cells were stained with Giemsa, a digital micrograph of the region was prepared and the stained area as a percent of total area in the scraped and open sub-regions was estimated by a thresholding procedure using IPLabs Spectrum (Scanalytics, Vienna, Virginia) software. Results in Table 1 represent data from 24 h after plating on coverslips treated with fibronectin (FN; $10 \mu\text{g ml}^{-1}$; ref. 26).

Examples of tubular network formation (associated with vasculogenic mimicry) could be observed following seeding of cell lines onto three-dimensional gels of polymerized Matrigel or Type 1 collagen (Collaborative Biochemical) as described (ref. 5; Table 1).

Table 1 lists results from high throughput screening for cell migration as the radial dispersion of cells from an initial confluent monolayer of 2,000 melanoma cells deposited within a 1.0 mm circular area on glass surfaces precoated with FN ($100 \mu\text{g ml}^{-1}$; refs 27, 28).

Selected cell lines were tested for their ability to invade a defined basement membrane matrix. Tumour cells (1×10^5) were seeded into the upper wells of the membrane invasion culture system (MICS) chamber²⁹ onto collagen/laminin/gelatin-coated (Sigma) polycarbonate membranes containing $10\text{-}\mu\text{m}$ pores (Osmonics, Livermore, California) containing $1 \times$ Mito+ Serum Extender (Becton Dickinson). After 24 h of incubation at 37°C , the cells that invaded each membrane were collected, stained and counted as described³⁰. Percent invasion was corrected for proliferation and calculated as (total number of invading cells/ total number of cells seeded) $\times 100$.

Received 12 January; accepted 16 May 2000.

- Hall, H. L., Miller, D. R., Rogers, J. D. & Bewerse, B. Update on the incidence and mortality from melanoma in the United States. *J. Am. Acad. Dermatol.* **40**, 35–42 (1999).
- Weyers, W., Euler, M., Diaz-Casajao, C., Schill, W. B. & Bonczkowitz, M. Classification of cutaneous malignant melanoma: a reassessment of histopathologic criteria for the distinction of different types. *Cancer* **86**, 288–299 (1999).
- Byers, H. R. & Bhawan, J. Pathologic parameters in the diagnosis and prognosis of primary cutaneous melanoma. *Hematol. Oncol. Clin. North Am.* **12**, 717–735 (1998).
- McMasters, K. M., Sondak, V. K., Lotze, M. T. & Ross, M. I. Recent advances in melanoma staging and therapy. *Ann. Surg. Oncol.* **6**, 467–475, (1999).
- Maniotis, A. J. et al. Vascular channel formation by human melanoma cells *in vivo* and *in vitro*: vasculogenic mimicry. *Am. J. Pathol.* **155**, 739–752 (1999).
- DeRisi, J. et al. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet.* **14**, 457–460 (1996).
- Khan, J. et al. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res.* **58**, 5009–5013 (1998).
- Perou, C. M. et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl Acad. Sci. USA* **96**, 9212–9217 (1999).
- Golub, T. R. et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286**, 531–537, (1999).
- Alizadeh, A. et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
- Bittner, M., Meltzer, P. & Trent, J. Data analysis and integration: of steps and arrows. *Nature Genet.* **22**, 213–215 (1999).
- Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
- Everitt, B. *Applied Multivariate Data Analysis*. (Oxford Univ. Press, New York, 1992)
- Ben-Dor, A., Shamir, R. & Yakhini, Z. Clustering gene expression patterns. *J. Comput. Biol.* **6**, 281–297 (1999).
- Adams, J. C. Characterization of cell-matrix adhesion requirements for the formation of fascin microspikes. *Mol. Biol. Cell* **8**, 2345–2363 (1997).
- Scott, G. & Liang, H. pp125FAK in human melanocytes and melanoma: expression and phosphorylation. *Exp. Cell Res.* **219**, 197–203 (1995).
- Janji, B., Melchior, C., Guon, V., Vallar, L. & Kieffer, N. Autocrine TGF- β -regulated expression of adhesion receptors and integrin-linked kinase in HT-144 melanoma cells correlates with their metastatic phenotype. *Int. J. Cancer* **83**, 255–262 (1999).
- Hicken, T. et al. Beta1 integrin expression in malignant melanoma predicts occult lymph node metastases. *Surgery* **118**, 669–673 (1995).
- Van Belle, P. et al. Progression-related expression of beta3 integrin in melanomas and nevi. *Hum. Pathol.* **30**, 562–567 (1999).
- Woods, A., Longley, R., Tumova, S. & Couchman, J. Syndecan-4 binding to the high affinity heparin-binding domain of fibronectin drives focal adhesion formation in fibroblasts. *Arch. Biochem. Biophys.* **374**, 66–72 (2000).
- Helge, C. et al. Interrelation of motility, cytoskeletal organization and gap junctional communication with invasiveness of melanocytic cells *in vitro*. *Invasion Metastasis* **17**, 26–41 (1997).
- Maug, K., Easty, D. J., Hill, S. & Bennett, D. Requirement for focal adhesion kinase in tumor cell adhesion. *Oncogene* **18**, 6824–6828 (1999).
- Silletti, S., Paku, S. & Raz, A. Autocrine motility factor and the extracellular matrix. I. Coordinate regulation of melanoma cell adhesion, spreading and migration involves focal contact reorganization. *Int. J. Cancer* **76**, 120–128 (1998).
- Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P. & Trent, J. Expression profiling using cDNA microarrays. *Nature Genet.* **21**, 10–14, (1999).
- Khan, J., Bittner, M., Chen, Y., Meltzer, P. & Trent, J. DNA Microarray technology: the anticipated impact on the study of human disease. *Biochim. Biophys. Acta* **1423**, 17–28 (1999).
- Tamura, M. et al. Inhibition of cell migration, spreading, and focal adhesions by tumor suppressor PTEN. *Science* **280**, 1614–1617 (1998).
- Berens, M., Rief, M., Loo, M. & Giese, A. The role of extracellular matrix in human astrocytoma migration and proliferation studied in a microliter scale assay. *Clin. Exp. Metastasis* **12**, 405–415 (1994).
- Giese, A., Loo, M., Norman, S., Treasurywala, S. & Berens, M. Contrasting migratory response of astrocytoma cells to tenascin mediated by different integrins. *J. Cell Sci.* **109**, 2161–8 (1996).
- Hendrix, M., Sefter, E., Sefter, R. & Fidler, I. A simple quantitative assay for studying the invasive potential of high and low human metastatic variants. *Cancer Lett.* **38**, 137–147 (1987).
- Hendrix, M., Sefter, E., Chu, Y., Trevor, K. & Sefter, R. Role of intermediate filaments in migration, invasion and metastasis. *Cancer Metastasis Rev.* **15**, 507–525 (1996).

Supplementary information is available on Nature's World-Wide Web site (<http://nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

Grant support from NIH (M.E.B., M.H.) and the Australian National Health and Medical Research Council (N.H.). We thank D. Edwards for access to instrumentation used in high throughput cell migration screening; A. Cress, K. Yamada and P. Schwartzberg for discussions on cell motility and assay design; and J. Pe'er, R. Folberg, K. Daniels and J. Kan-Mitchell for donation of uveal melanoma cell lines described in ref. 5.

Correspondence and requests for materials should be addressed to J.T. (jtrent@nih.gov) or M.B. (mibittner@nhgri.nih.gov).

A chromatin remodelling complex involved in transcription and DNA processing

Xuetong Shen*, Gaku Mizuguchi†, Ali Hamiche* & Carl Wu*

* Laboratory of Molecular Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-4255, USA

† Kyoto University Graduate School of Biostudies, Sakyo-ku, Kyoto, 606-8502, Japan

The packaging of the eukaryotic genome in chromatin presents barriers that restrict the access of enzymes that process DNA^{1,2}. To overcome these barriers, cells possess a number of multi-protein, ATP-dependent chromatin remodelling complexes, each containing an ATPase subunit from the SNF2/SWI2 superfamily^{3,4}. Chromatin remodelling complexes function by increasing nucleosome mobility and are clearly implicated in transcription⁵⁻⁷. Here we have analysed SNF2/SWI2- and ISWI-related proteins to identify remodelling complexes that potentially assist other DNA transactions. We purified a complex from *Saccharomyces cerevisiae* that contains the Ino80 ATPase⁸. The INO80 complex contains about 12 polypeptides including two proteins related to the bacterial RuvB DNA helicase⁹⁻¹¹, which catalyses branch migration of Holliday junctions. The purified complex remodels chromatin, facilitates transcription *in vitro* and displays 3' to 5' DNA helicase activity. Mutants of *ino80* show hypersensitivity to agents that cause DNA damage, in addition to defects in transcription⁸. These results indicate that chromatin remodelling driven by the Ino80 ATPase may be connected to transcription as well as DNA damage repair.

We searched the *Saccharomyces* genome database for genes that were highly related to *Drosophila* ISWI, which codes for the ATPase subunit of the nucleosome remodelling factor NURF¹². We found *SWI2/SNF2*, *STH1*, *ISWI*, *ISW2*, *CHD1* and the three open reading frames (ORFs) *YFR038W*, *YGL150C* and *YDR334W*. Ebbert *et al.*⁸ identified *YGL150C* in a genetic screen for mutants affecting inositol biosynthesis. Previously we had called it *ARI1* (for ATPase Related to

ISWI); hereafter, we use the *INO80* (ref. 8) nomenclature for this gene. Typical of SNF2/SWI2 family proteins, the conserved ATPase domain of Ino80 comprises a large proportion of the coding region. Beyond the ATPase domain, Ino80 does not possess the SANT domain found in Isw1 and Isw2 (ref. 13); however, sequence comparison of *INO80* and its counterparts from human (*hINO80*) and *Drosophila* (*dINO80*) reveals two additional conserved regions, the TELY motif at the amino terminus and the GTIE motif at the carboxy terminus.

To analyse Ino80 biochemically, we engineered a double-Flag tag at the C terminus of the chromosomal *INO80* gene by homologous recombination. The behaviour of the strain carrying INO80-Flag was identical to wild type under all conditions tested. Immunostaining showed that Ino80-Flag is localized unevenly in the nucleus and is associated with the chromosomes of dividing cells (data not shown). Gel-filtration chromatography of INO80-Flag cell extracts shows the bulk of Ino80 in a high relative molecular mass (M_r) fraction between 1,000K and 1,500K (Fig. 1a), with no detectable material in smaller fractions. Therefore most, if not all, of Ino80 is present in a large complex. The high M_r of native Ino80 is partly consistent with a previous study, which found both high and low M_r fractions⁸; the presence of a low M_r species, presumably uncomplexed or partially complexed Ino80, may be an artefact of overexpression.

We purified the complex of proteins associated with Ino80, the 'INO80 complex' (INO80.com), by immunoprecipitation with anti-Flag agarose. The purified INO80 complex contains 12 principal polypeptides, most of which, with notable exceptions, are present at roughly equivalent stoichiometry (Fig. 1b). We sequenced the peptides of individual protein bands by microcapillary reverse-phase high performance liquid chromatography nano-electrospray ion trap mass spectrometry. Data identifying seven proteins in the complex are given in Supplementary Information. Actin (Act1) and three actin-related proteins, Arp4, Arp5 and Arp8, are associated with the complex in addition to Ino80. β -Actin is a functional component of the mammalian BAF complex¹⁴, and Arp7 and Arp9 are shared by the yeast SWI/SNF and RSC complexes¹⁵. Actin and Arp4 are also contained in the yeast NuA4 histone acetyltransferase complex¹⁶. Two other intensely staining polypeptides of the INO80 complex, Rvb1 and Rvb2, are encoded by the ORFs *YDR190C* and

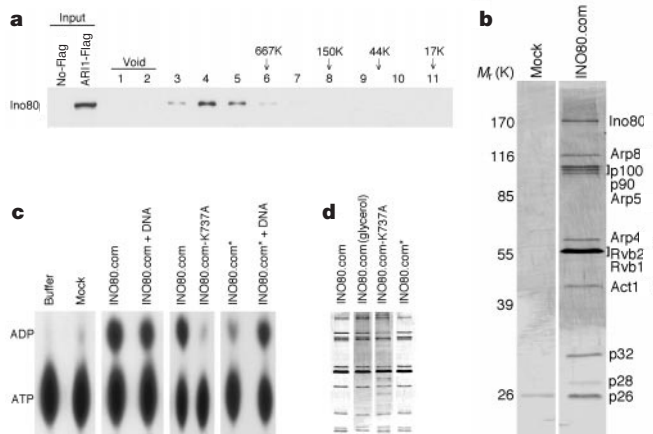


Figure 1 Purification and characterization of the INO80 complex. **a**, Superose-6 chromatography of whole-cell extract from INO80-Flag cells. Fractions containing Ino80 were detected by western blotting with anti-Flag antibody. **b**, SDS-PAGE and silver staining showing immunoaffinity-purified INO80 complex (INO80.com). Mock purification eluted only the IgG light chain. **c**, Thin-layer chromatography analysis showing ATPase activity of INO80 complexes INO80.com, INO80.com-K737A and INO80.com* (see text for details). DNA was double-stranded Φ X174 RF DNA. **d**, SDS-PAGE and silver staining showing composition of INO80 complexes.

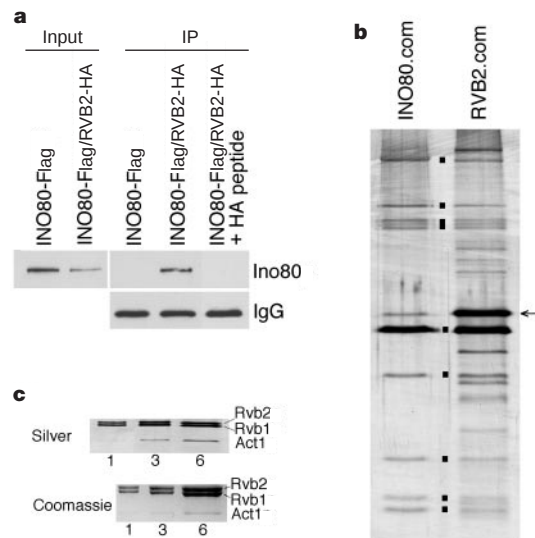


Figure 2 Rvb1 and Rvb2 are subunits of the INO80 complex. **a**, Western blots showing immunoprecipitation of Ino80 with Rvb2-HA using anti-HA antibody and detection with anti-Flag antibody. **b**, SDS-PAGE and silver stain showing the RVB2 complexes (RVB2.com). Filled dots indicating bands shared with INO80.com. Note the extra bands in RVB2.com. Arrow indicates Rvb2-Flag. **c**, SDS-PAGE showing stoichiometry of Rvb1 and Rvb2 in the INO80 complex. Increasing amounts of INO80 complex (1 $\times$, 3 $\times$ and 6 $\times$) were analysed by SDS-PAGE.

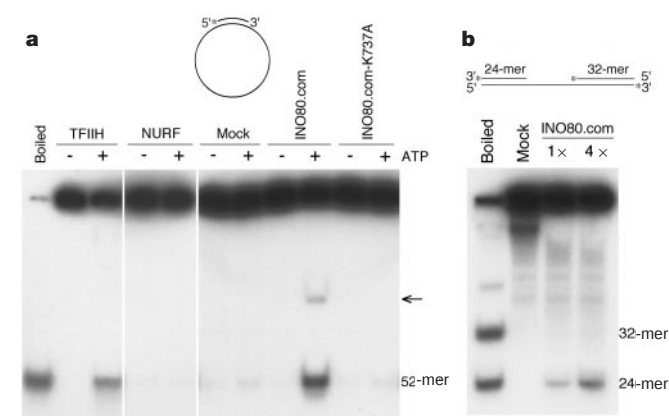


Figure 3 INO80 complex has 3' to 5' specific helicase activity. **a**, Native PAGE (10%) showing the displacement of a 32 P-labelled 52-mer annealed to circular single-stranded DNA on incubation with the indicated reagents. (INO80.com also produces a weak band (arrow), which may be a metastable intermediate.) **b**, Native PAGE (10%) showing the displacement of the 32 P-labelled 24-mer or 32-mer. 1× and 4× refer to concentrations of INO80 complex. All reactions contained ATP. Structures of substrates are shown on top and the 32 P-label is indicated by an asterisk.

YPL235W, respectively. Rvb1 and Rvb2 were previously identified as 'RuvB-like' proteins (scRUVBL1/scTIP49a and scRUVBL2/scTIP49b) sharing homology to bacterial RuvB^{9,10}, the Holliday junction DNA helicase¹⁷; both *RVB1* and *RVB2* genes are essential in yeast^{9,11}. The first RuvB-like protein (TIP49/rTIP49a) was identified in rat¹⁰, and a pair of RuvB-like proteins (hTIP49a/RUVBL1/TIP49 and hTIP49b/TIP48) are found in human and are associated in large complexes^{9,11,18,19}.

Like other ATP-driven chromatin remodelling complexes, the INO80 complex has ATPase activity (Fig. 1c, INO80.com). The ATPase activity was intrinsically high, however, and was not further stimulated by exogenous DNA (Fig. 1c, INO80.com+DNA). We ascribe the bulk of this ATPase activity to Ino80, as the complex immunopurified from a strain carrying a K737A substitution in the ATP-binding site retained the same subunit composition (Fig. 1d, INO80.com-K737A), but showed substantially reduced activity (Fig. 1c, INO80.com-K737A). The remaining activity (~5%) may be derived from other subunits of the complex, or from the K737A protein. The inability to stimulate ATPase activity by exogenous DNA might be caused by the presence of endogenous DNA associated with the INO80 complex. We detected DNA of heterogeneous size in the purified preparation, and effectively removed the nucleic acid by extensive digestion with DNase I (data not shown). DNase I digestion had no observable effects on the polypeptide composition of the INO80 complex (Fig. 1d, INO80.com*), but the intrinsic ATPase activity of the INO80 complex was substantially reduced (Fig. 1c, INO80.com*). We could then stimulate the ATPase activity of the DNase I-treated complex with exogenous DNA (Fig. 1c, INO80.com*+DNA), or with nucleosomes (data not shown). Thus, like the related remodelling complexes SWI/SNF²⁰ and RSC²¹, the INO80 complex shows DNA-dependent ATPase activity.

Among the polypeptides associated with Ino80, Rvb1 and Rvb2 are of special interest. To confirm that these proteins are subunits of the INO80 complex, we expressed Rvb2 from a plasmid carrying *RVB2* tagged with a haemagglutinin A (HA) epitope in INO80-Flag cells. Immunoprecipitation with anti-HA antibody revealed a specific interaction with Ino80 (Fig. 2a), strongly suggesting that this RuvB-like protein (and probably Rvb1 as well) is a true component of the INO80 complex. To assess whether Ino80 forms one complex or several complexes each containing a different subset of the 11 associated polypeptides, we analysed the proteins associating with Rvb2 in a wild-type strain transformed with a

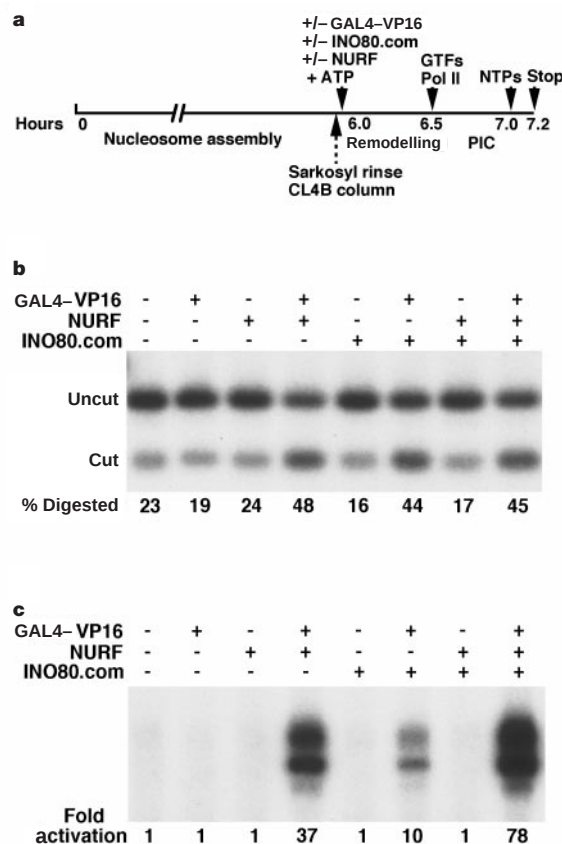


Figure 4 INO80 complex remodels and promotes transcription from a chromatin template. **a**, Protocol. **b**, Southern blots showing that both NURF and the INO80 complex enhance digestion of a *PstI*-*ClaI* fragment (uncut) by *Bam*HI (cut). **c**, Transcription analysis of remodelled chromatin templates. Primer extension of transcripts from the E4 promoter (+99 to +72 primer). ATP was present in all reactions and was needed for the remodelling step.

RVB2-Flag construct. Immunoprecipitation of Rvb2-Flag revealed that 11 out of the 12 proteins of the INO80 complex co-purify with Rvb2 (Fig. 2b, filled dots). The twelfth polypeptide, Arp4, may be obscured by the co-migrating Rvb2-Flag protein band (Fig. 2b, arrow). Moreover, glycerol gradient centrifugation of the original, immunopurified INO80 complex (from INO80-Flag cells) showed that all of the 12 polypeptides sedimented together in a high M_r fraction (Fig. 1d, INO80.com (glycerol)). These data suggest that Ino80 is contained in a single protein complex of about 12 distinct proteins. In addition, a number of new proteins co-purified with Rvb2-Flag (Fig. 2b). The presence of these polypeptides, which are evidently not part of the INO80 complex, may expand the functions of Rvb2 and explain why, in contrast to *ino80* mutants, *rvb2* mutants are lethal¹¹.

The staining intensity of Rvb1 and Rvb2 protein bands in the INO80 complex is unusually high in relation to other polypeptides (Fig. 1b). Analysis on a higher resolution gel revealed a closely spaced doublet band corresponding to the 50K Rvb1 and 52K Rvb2 polypeptides (Fig. 2c). We determined the stoichiometry of the RuvB-like proteins by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and densitometry after staining with silver or Coomassie blue (Fig. 2c). The molar ratios for Rvb1 and Rvb2 relative to Act1 are 6.5 and 5.6 (silver), and 6.3 and 6.7 (Coomassie), respectively, whereas the ratios relative to Ino80 are 5.3 and 5.6 (Coomassie). To a first approximation, both Rvb1 and Rvb2 show 6:1 stoichiometry compared with other polypeptides in the INO80 complex. The correspondence between these values and the known double hexamer composition of bacterial RuvB is striking¹⁷.

Given that bacterial RuvB¹⁷ is an ATP-dependent DNA helicase, we investigated whether the INO80 complex containing Rvb1 and Rvb2 subunits has helicase activity. For these experiments and the *in vitro* transcription studies below, we used the INO80 complex without DNase I treatment because of difficulty in eliminating residual DNase I after digestion. We found that the INO80 complex was able to displace a radiolabelled 52-nucleotide DNA strand from a duplex formed by annealing the oligonucleotide to single-stranded Φ X174 DNA (Fig. 3a). Primer displacement was ATP-dependent, as noted for other helicases such as TFIIH. In contrast, *Drosophila* NURF showed no evidence of helicase activity, despite its potency in chromatin remodelling. The Ino80-K737A mutant complex failed to show DNA helicase activity (Fig. 3a), indicating that Rvb1 and Rvb2 may be functionally coupled to the Ino80 ATPase. (The ability of RuvB-like proteins to function independently as helicases is a subject of debate^{9,11,19}.) To examine the directionality of helicase action, we used a substrate consisting of 24-mer and 32-mer oligonucleotides annealed to the corresponding ends of linear, single-stranded Φ X174 DNA (Fig. 3b, top). The INO80 complex was able to displace only the 24-mer, indicating a 3' to 5' helicase activity (Fig. 3b). The directional helicase activity of the INO80 complex was also ATP dependent (data not shown).

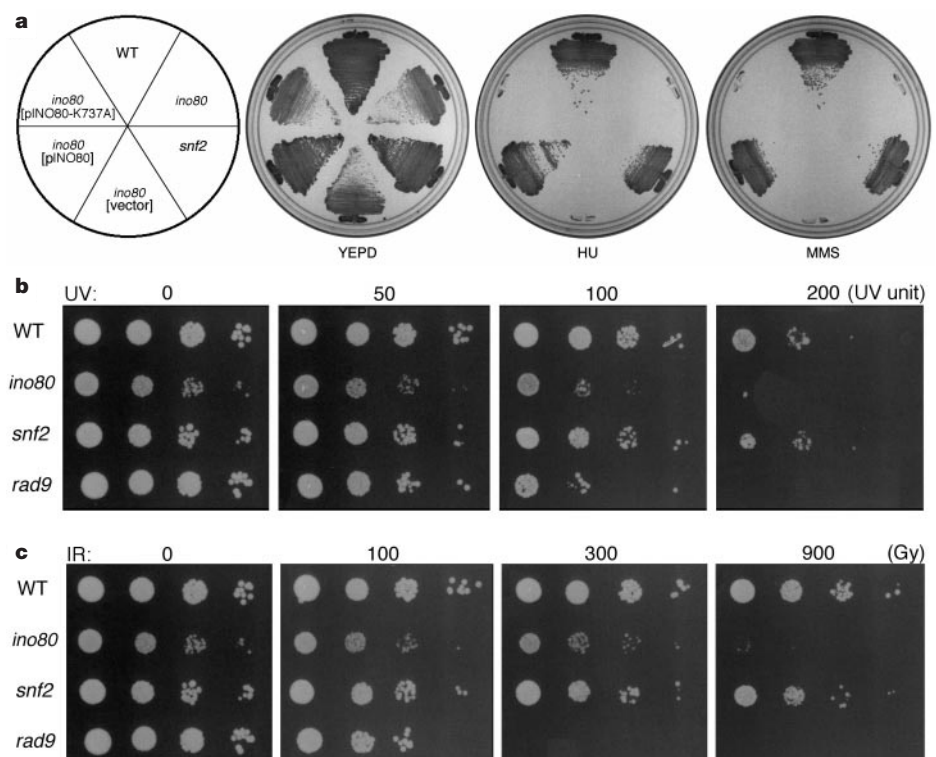
In previous work, *ino80* mutants exhibited reduced transcription of a number of reporter constructs⁸. In concurrence, we found reduced RNA levels for the endogenous *INO1*, *PHO5* and *Ty1* genes in *ino80* mutant cells but not for *INO2*, *PHO8* and *ACT1*, and normal activation but retarded attenuation of *HSP26* (data not shown). We also investigated the ability of the INO80 complex to affect transcription *in vitro* by reconstituting plasmid DNA carrying five GAL4-binding sites upstream of the adenovirus E4 core promoter into chromatin²². We incubated the chromatin template with a saturating amount of the GAL4-VP16 transactivator, INO80 complex and ATP (Fig. 4a), and measured remodelling activity by the change in accessibility to *Bam*HI digestion at a site upstream (−46) of the E4 TATA box. The INO80 complex displayed activity (44% digestion) comparable to the remodelling activity of NURF (48% digestion), whereas control reactions lacking the GAL4-VP16

transactivator, the INO80 complex or both, gave substantially lower cleavage (16–23% digestion) (Fig. 4b). We reproducibly observed a 10-fold transcriptional activation of chromatin remodelled by a saturating amount of the INO80 complex (Fig. 4a, c), which is lower than the 37-fold activation typically observed for NURF (Fig. 4c). Curiously, the combined actions of NURF and the INO80 complex, while creating no greater accessibility (45% digestion), resulted in 78-fold activation; the synergistic activation may reflect differences in the respective remodelling mechanisms.

To explore further the physiological functions of *INO80*, we investigated the phenotype of the *ino80* null mutant, and found it to be sensitive to hydroxyurea (100 mM), and the alkylating agent methyl methanesulphonate (MMS; 0.025%) (Fig. 5a). Both agents affect DNA metabolism: hydroxyurea inhibits DNA synthesis by reducing the dNTP pool; and MMS induces single- and double-stranded breaks in DNA. The *ino80* mutant also showed sensitivity to ultraviolet and ionizing radiation, which cause DNA damage. By analysis of serially diluted cultures, the *ino80* mutant was as sensitive to ultraviolet as a *rad9* null mutant in the same genetic background; *RAD9* is a cell-cycle checkpoint gene required for induction of genes that repair DNA damage²³ (Fig. 5b). The *ino80* mutant also showed sensitivity to ionizing radiation, although the degree of sensitivity was intermediate when compared with the *rad9* mutant (Fig. 5c). In contrast, a *snf2* null mutant in the same genetic background was insensitive to identical levels of hydroxyurea and MMS, and insensitive to ultraviolet or ionizing radiation (Fig. 5). The sensitivity to hydroxyurea, MMS, ultraviolet and ionizing radiation is due to the loss of *INO80*, as the null phenotype can be rescued with a plasmid carrying the wild-type gene (Fig. 5a, pINO80; and data not shown). Moreover, a K737A substitution in the ATP-binding motif abolished the ability to rescue the mutant phenotypes, indicating the importance of ATP-binding to *INO80* function *in vivo* (Fig. 5a, pINO80-K737A; and data not shown). These results indicate that *INO80* is involved, directly or indirectly, in the processing of DNA damage.

The INO80 complex might function indirectly in DNA repair by facilitating transcription of genes induced by DNA damaging

Figure 5 Analysis of *ino80* mutant phenotypes. **a**, Strains used are wild type (WT), *ino80* null mutant (*ino80*), *snf2* null mutant (*snf2*), *ino80* with empty vector (*ino80*[Vector]), *ino80* with vector containing the wild-type *ino80* gene (*ino80*[pINO80]) and *ino80* with vector containing the K737A mutant (*ino80*[pINO80-K737A]). HU, hydroxyurea; MMS, methyl methanesulphonate. **b**, Serial dilutions of cells from WT, *ino80*, *snf2* and *rad9* null mutant (*rad9*) strains were given increasing doses of ultraviolet (UV) as indicated. **c**, Similar to **b**, strains were subjected to ionizing radiation (IR) at the indicated doses.



agents; however, we have found normal induction of the ribonucleotide reductase genes *RNR1* and *RNR3* (which are the downstream targets of the DNA damage checkpoint gene *MEC1*; ref. 24) after treatment of *ino80* cells with hydroxyurea (data not shown). Alternatively, chromatin remodelling by the INO80 complex might directly facilitate an aspect of DNA metabolism involving replication, recombination or repair, processes that are intimately connected²⁵. The INO80 complex might either promote site recognition by the replication or repair machinery, or assist progression of the machinery through chromatin. In *Escherichia coli*, the RuvAB complex binds to Holliday junctions and promotes branch migration, the motive force being provided by the helicase activity of the two hexameric RuvB rings¹⁷. The RuvAB complex also has a role in the generation and prevention of double-stranded breaks at arrested replication forks through its actions on potential Holliday junction intermediates²⁶. The yeast Rvb1 and Rvb2 hexamers in the INO80 complex might function similarly, but with the added coupling to chromatin remodelling proteins as an adaptation to the chromatin environment. Our results, together with the recent findings of human RuvB-like proteins being associated with c-Myc¹⁸ and being contained in the multisubunit TIP60 HAT complex¹⁹, underscore the linkage of this class of DNA helicases with diverse aspects of chromatin metabolism. □

Methods

Bioinformatics and gene cloning

We carried out the yeast gene search in the *Saccharomyces* genome database. A *hINO80* complementary DNA was cloned by filling a gap between the two expressed sequence tags using PCR, and was found to be nearly identical to a complementary DNA encoding the uncharacterized KIAA1259 protein. *dINO80* was predicted from genomic sequences.

Manipulation of yeast

All strains (see Supplementary Information) are in the S288C background. The chromosomal *INO80* gene was tagged at the extreme C terminus by integrating a pRS406 plasmid containing the *XhoI*–*HindIII* fragment of *INO80* in which a double Flag sequence was introduced right before the stop codon (pRS406–*INO80*/2Flag). The URA3 marker was 'popped out' through homologous recombination. The tagged *INO80*–Flag strain was confirmed by Southern blotting, PCR and sequencing. The *RVB2* ORF was tagged with HA at the C terminus in a low-copy plasmid pYX112 (pRVB2–HA). Similarly, the *INO80* ORF was cloned into pYX112 (pINO80) and a single amino-acid substitution (K737A) was then made (pINO80–K737A). For purification of *INO80.com*–K737A and *RVB2.com*, we cloned fragments of ORFs with native promoters (to –500 and to –1,000, respectively) into a modified pRS416 plasmid, which contains a double Flag sequence. The *ino80* strain was created using the PCR deletion method and found to be viable; however, in a W303 strain background, *INO80* is essential and *ino80* temperature-sensitive mutants showed a G2/M arrest phenotype upon temperature shift (unpublished data). The *snf2* strain was a gift from F. Winston (Harvard University) and the *rad9* strain was from Research Genetics. Phenotypic analysis was done by subjecting cells to 100 mM hydroxyurea, 0.025% MMS, ultraviolet in a Stratalinker or ionizing radiation using a Cs-137 gamma ray source on YEYP plates. We incubated plates at 30 °C for 2–3 days and then scored.

Protein techniques

Four litres of *INO80*–Flag cells were grown to saturation for 2 days at 30 °C in YEYP. Whole-cell extracts were prepared essentially as described¹³. About 40 ml of extract in Buffer H-0.3 (ref. 13) was routinely obtained. Direct immuno-affinity purification of the *INO80* complex was done essentially as described¹³. To remove associated DNA, DNase I (50 µg) in buffer H-0.1 was added to beads after washing with buffer H-0.15 and allowed to digest for 30 min at 24 °C. The yield was roughly 10 µg protein complex per litre of cells. *INO80.com*–K737A and *RVB2.com* were similarly purified from epitope-tagged cell extracts. We carried out mock purification using extract made from wild-type cells. Gel-filtration chromatography was on a Superose-6 column in buffer H-0.3. 17–35% glycerol gradient sedimentation was in buffer H-0.1. Silver and Coomassie stained gels were scanned and quantified with NIH Image software. Peptide sequencing was done at the Harvard Microchemistry Facility on a Finnigan LCQ quadrupole ion trap mass spectrometer.

In vitro activity assays

The ATPase and restriction accessibility assays were carried out as described^{13,27}. Chromatin transcription was done essentially as described²², except that GAL4–VP16 was used as transactivator and that purified *Drosophila* RNA polymerase II and purified and recombinant general transcription factors were used instead of nuclear extract. For the helicase assay, a ³²P-labelled 52-mer was annealed to single-stranded ΦX174 DNA at position 702–753 (ref. 9). Reactions were carried out in 10 µl for 30 min at 30 °C and

included 20 ng of complexes (1×). TFIID²⁸ and NURF²⁹ were purified as described. The substrate for directional helicase activity was prepared essentially as described³⁰ except that dGTP was used to blunt ends after cutting with *HapII* of the helicase substrate (see above) and filling-in with ³²P-dCTP, resulting in the generation of a 24-mer and a 32-mer.

Received 20 April; accepted 20 June 2000.

- Kornberg, R. D. & Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285–294 (1999).
- Luger, K. & Richmond, T. J. DNA binding within the nucleosome core. *Curr. Opin. Struct. Biol.* **8**, 33–40 (1998).
- Eisen, J. A., Sweder, K. S. & Hanawalt, P. C. Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Res.* **23**, 2715–2723 (1995).
- Peterson, C. L. Multiple SWItches to turn on chromatin? *Curr. Opin. Genet. Dev.* **6**, 171–175 (1996).
- Armstrong, J. A. & Emerson, B. M. Transcription of chromatin: these are complex times. *Curr. Opin. Genet. Dev.* **8**, 165–172 (1998).
- Kadonaga, J. T. Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell* **92**, 307–313 (1998).
- Workman, J. L. & Kingston, R. E. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu. Rev. Biochem.* **67**, 545–579 (1998).
- Ebbert, R., Birkmann, A. & Schüller, H. J. The product of the SNF2/SWI2 paralogue INO80 of *Saccharomyces cerevisiae* required for efficient expression of various yeast structural genes is part of a high-molecular-weight protein complex. *Mol. Microbiol.* **32**, 741–751 (1999).
- Qiu, X. B. *et al.* An eukaryotic RuvB-like protein (RUVBL1) essential for growth. *J. Biol. Chem.* **273**, 27786–27793 (1998).
- Kanemaki, M. *et al.* Molecular cloning of a rat 49-kDa TBP-interacting protein (TIP49) that is highly homologous to the bacterial RuvB. *Biochem. Biophys. Res. Commun.* **235**, 64–68 (1997).
- Kanemaki, M. *et al.* TIP49b, a new RuvB-like DNA helicase, is included in a complex together with another RuvB-like DNA helicase, TIP49a. *J. Biol. Chem.* **274**, 22437–22444 (1999).
- Tsukiyama, T., Daniel, C., Tamkun, J. & Wu, C. ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodelling factor. *Cell* **83**, 1021–1026 (1995).
- Tsukiyama, T., Palmer, J., Landel, C. C., Shiloach, J. & Wu, C. Characterization of the imitation switch subfamily of ATP-dependent chromatin-remodelling factors in *Saccharomyces cerevisiae*. *Genes Dev.* **13**, 686–697 (1999).
- Zhao, K. *et al.* Rapid and phosphoinositide-dependent binding of the SWI/SNF-like BAF complex to chromatin after T lymphocyte receptor signaling. *Cell* **95**, 625–636 (1998).
- Cairns, B. R., Erdjument-Bromage, H., Tempst, P., Winston, F. & Kornberg, R. D. Two actin-related proteins are shared functional components of the chromatin-remodelling complexes RSC and SWI/SNF. *Mol. Cell* **2**, 639–651 (1998).
- Galarneau, L. *et al.* Multiple links between the NuA4 histone acetyltransferase complex and epigenetic control of transcription. *Mol. Cell* **5**, 927–937 (2000).
- West, S. C. Processing of recombination intermediates by the RuvABC proteins. *Annu. Rev. Genet.* **31**, 213–244 (1997).
- Wood, M. A., McMahon, S. B. & Cole, M. D. An ATPase/helicase complex is an essential cofactor for oncogenic transformation by c-Myc. *Mol. Cell* **5**, 321–330 (2000).
- Ikura, T. *et al.* Link of TIP60 histone acetylase to DNA repair and apoptosis. *Nature* (submitted).
- Côté, J., Quinn, J., Workman, J. L. & Peterson, C. L. Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science* **265**, 53–60 (1994).
- Cairns, B. R. *et al.* RSC, an essential, abundant chromatin-remodelling complex. *Cell* **87**, 1249–1260 (1996).
- Mizuguchi, G., Tsukiyama, T., Wisniewski, J. & Wu, C. Role of nucleosome remodelling factor NURF in transcriptional activation of chromatin. *Mol. Cell* **1**, 141–150 (1997).
- Weinert, T. A. & Hartwell, L. H. Characterization of RAD9 of *Saccharomyces cerevisiae* and evidence that its function acts posttranslationally in cell cycle arrest after DNA damage. *Mol. Cell. Biol.* **10**, 6554–6564 (1990).
- Huang, M. & Elledge, S. J. Identification of RNR4, encoding a second essential small subunit of ribonucleotide reductase in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **17**, 6105–6113 (1997).
- Haber, J. E. DNA recombination: the replication connection. *Trends Biochem. Sci.* **24**, 271–275 (1999).
- Seigneur, M., Bidnenko, V., Ehrlich, S. D. & Michel, B. RuvAB acts at arrested replication forks. *Cell* **95**, 419–430 (1998).
- Gdula, D. A., Sandaltzopoulos, R., Tsukiyama, T., Ossipow, V. & Wu, C. Inorganic pyrophosphatase is a component of the *Drosophila* nucleosome remodelling factor complex. *Genes Dev.* **12**, 3206–3216 (1998).
- Hansen, S. K. & Tjian, R. TAFs and TFIID mediate differential utilization of the tandem Adh promoters. *Cell* **82**, 565–575 (1995).
- Tsukiyama, T. & Wu, C. Purification and properties of an ATP-dependent nucleosome remodelling factor. *Cell* **83**, 1011–1020 (1995).
- Park, J. S., Choi, E., Lee, S. H., Lee, C. & Seo, Y. S. A DNA helicase from *Schizosaccharomyces pombe* stimulated by single-stranded DNA-binding protein at low ATP concentration. *J. Biol. Chem.* **272**, 18910–18919 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank T. Tsukiyama and the CSHL yeast genetics course 1998 for expert training, and members of our lab for help and discussion. We also thank Y. Nakatani, J. Pradel and J. Cote for sharing results; W. S. Lane for peptide sequencing; F. Winston and D. E. Gottschling for yeast strains; and M. Lichten and G. Storz for helpful comments on the manuscript. X.S. was supported by the American Cancer Society, and A.H. was supported by the Human Frontiers of Science Program. This work was supported by the Intramural Research Program of the National Cancer Institute.

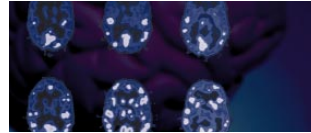
Correspondence and requests for materials should be addressed to C.W. (e-mail: carlwu@helix.nih.gov).



Head start

Oregon and Boston stake their claims in neuroscience

p545



Tackling brain tumours

A new centre at Dana-Farber aims to crack the code behind brain cancer

p546

Collaborations prepare to untangle the circuitry of the brain

Scientists from many disciplines are putting their heads together in an attempt to understand how the brain works. Diane Gershon listens in.

Modern neuroscience is at the crossroads of many scientific disciplines and, as such, would clearly benefit from a more multidisciplinary approach to problem solving. Two recent initiatives, in Oregon and Boston, show that there is much to be gained by removing the traditional barriers between disciplines and building multidisciplinary research programmes that draw on a broad base of expertise.

This year the Neurological Sciences Institute (NSI) at Oregon Health Sciences University (OHSU) in Portland, Oregon, embarked on an expansion programme. As part of this, it will increase the breadth and scope of its neuroscience research, as well as the range of neurological disorders being studied.

Celebrating its 25th anniversary this year, the NSI only became part of OHSU in April 1998. What makes it different from other institutes, says Paul Cordo, the NSI's director and senior scientist, "is that we don't specialize: we have everything from molecular biology up to human perception being worked on". Moreover, research at the institute runs the gamut from basic to applied to clinical. Central to its mission is the translation of research into advances in healthcare.

This ability to bridge the gap between basic and clinical science "is one of the things that we bring to the [OHSU] medical school", says Cordo. Since joining OHSU, the NSI has already formed close to 20 collaborative research projects with other departments, about half of which are clinically based. Cordo's own research interests centre on how the human nervous system coordinates voluntary movement and how proprioception — the mostly subconscious sense of body position and movement — contributes to coordination. He has an ongoing clinical collaboration to develop new approaches to stroke rehabilitation.

In December, the NSI will move to a building on the university's new west campus in Hillsboro. This is 15 miles from the main site and will provide much-needed expansion space for OHSU. The NSI building, together



Grand plans: the Neurological Sciences Institute covers both basic research and clinical application.

There are no preconceived ideas about who to hire other than they do outstanding research.

with a new Vaccine and Gene Therapy Institute, represents the first phase of university expansion on the 240-acre west campus site.

At full capacity, the NSI's 60,000-square-foot building will house 22 research labs and one theoretical research group. The institute recently hired seven researchers as part of a first round of staff expansion, bringing the number of NSI investigators to 17. Cordo hopes to fill the remaining six positions over the next year as "empty labs don't generate any science or income", he says.

When asked what he is looking for in new recruits, Cordo says he has "no preconceived notions about who to hire other than that they do outstanding research". But as about 90% of the NSI's budget comes from external research grants, he says he is looking for peo-

ple who are committed to research, who will be self-supporting and capable of sustaining an independent research programme over their entire career.

Members of the NSI are not tenured. And, although this may send some researchers into a cold sweat, Cordo says the NSI has a good track record when it comes to supporting itself with outside grants. In its 25-year existence, he says he knows of only two people who left because they could not secure funding. Researchers going through a lean period can get 'bridge funding' from the institute, which is guaranteed for two years and may be extended at the director's discretion.

Unlike most university departments, "we have no teaching or administrative requirements for our staff", says Cordo. But, he says, most researchers do teach on a voluntary basis as a way of gaining access to graduate students.

Whereas NSI is interesting because of the breadth of research covered, the new institute for brain research at the Massachusetts Institute of Technology (MIT) will have a keen focus on how the brain works as a central basic biological and physical problem. The research will be very fundamental, says the institute's founding director, Nobel laureate Phillip Sharp. And, unlike the NSI,



Sharp (left) and Lore Harp McGovern (centre) at the official creation of the McGovern Institute.

▶ Sharp does not anticipate that the new institute will have a disease focus — although behavioural research will feature.

Creation of the McGovern Institute for Brain Research, as it will be called, has been made possible by a sizeable gift from philanthropists Patrick McGovern, founder and chairman of the International Data Group (IDG), and his wife, the entrepreneur Lore Harp McGovern. The gift, which was announced in February, is expected to total about \$350 million over the next 20 years.

Sharp, who is head of MIT's department of biology, sees his new role as building a



Sharp: will focus on the fundamentals.

critical mass in the institute while keeping the research focused on the brain as a functioning biological entity. "That doesn't mean it necessarily has to be molecular, cellular or genetic," he says, but it should aim to deepen researchers' understanding of how the brain stores and processes information and makes decisions. By keeping the institute's sights firmly fixed on the fundamental problems, Sharp anticipates that the research will shed light on diseases such as Alzheimer's and schizophrenia.

The full power of the institute will not be realized for at least ten years or so, Sharp says. "That gives you a chance to mould, as the science evolves, the focus of the institute so you don't have to make all your bets right up front." At full strength, it will be a sizeable operation, involving 300–400 people, although it will be centred around 16 McGovern investigators, ten of whom will be new recruits. A new building is planned for a site in the middle of campus, providing 94,000 square feet, although the new home will not be ready until 2004.

Even before the McGovern gift was announced, neuroscience had been singled out as a strategic priority for MIT and its School of Sciences. The brain and cognitive sciences department, for example, is expanding and the brain and cognitive sciences major is one of the fastest growing at MIT. There is

also the Center for Learning and Memory and the newly created Martinos Center for Functional and Structural Biomedical Imaging. Added to this, there are various pockets of research on campus with applications to neuroscience in areas such as artificial intelligence and computational science.

The stated objective of the new institute is to be inter-departmental. McGovern investigators will all have a departmental 'home' and, as such, will have to fulfil the usual departmental obligations, such as teaching, but Sharp

says: "The institute will have the resources and space that will allow the intellectual issues to drive recruitment and development of programmes and not departmental issues."

For Sharp, who was awarded a Nobel Prize in 1993 for his discovery of split genes and RNA splicing, being director of the McGovern brain institute will add a new dimension to his career. But he will continue to pursue his own research interests. "I'm not holding any great aspirations that I will become a great neuroscientist," he quips. ■

Mahoney centre set to tackle brain cancer head on

A rather perverse scenario is behind an multi-institutional research initiative set up by the Dana-Farber Cancer Institute (DFCI) in Boston. Launched in May, the David Mahoney Center for Neuro-Oncology will focus on some of the most intractable diseases of the brain.

This move was sparked by the distressing figures on primary tumours of the central nervous system (CNS). These account for less than 2% of all new reported cases of cancer in the United States each year, but "the majority of these cancers are glioblastomas, and so are — to all intents and purposes — incurable", says Charles Stiles, chairman of the department of cancer biology at the DFCI. In these cases, the median time from diagnosis to death is about five months, he says, so, although they are quite rare, when it comes to the causes of cancer-related deaths, "they're right up there with the heavy hitters", such as breast, prostate, lung and colon cancer.

Primary tumours of the CNS are the third most common cause of cancer-related death among men in the 15–34 age group, and they are the most common form of solid malignancy in children. Moreover, primary cancers of neuro-ectodermal origin are one of the leading causes of death in children.

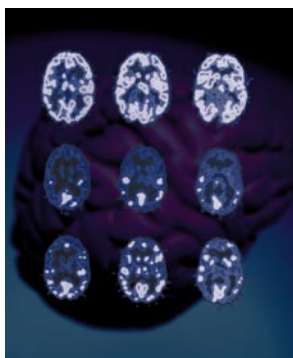
Stiles says the DFCI's board of trustees concluded that this was exactly the sort of problem that a cancer institute in a major metropolitan area ought to be throwing its weight behind. "If you want to work on brain cancer, you have to work where you can get the caseload," he says.

The new centre came about after the Charles A. Dana Foundation pledged \$7.4 million in support. David Mahoney, the cen-

tre's namesake, was a trustee of the DFCI and served as chairman and chief executive officer of the foundation until his death in May.

The new programme in neuro-oncology will draw on relevant expertise at the DFCI, Harvard Medical School — most notably the department of neurobiology now chaired by Carla Shatz — the School of Public Health, and all of the other Harvard-affiliated teaching hospitals. Stiles, who is deputy principal investigator of the programme (DFCI president David Nathan is its principal investigator), says it provides a "small concrete example" of a much broader ongoing effort to coordinate cancer research activities across these institutions.

The neuro-oncology programme will provide support for the development of certain core facilities, such as a brain tumour tissue bank, neuropathology, bioinformatics, as well as the purchase of shared equipment. But contained within the programme is a project to clone and characterize genes that direct the development of the normal human brain. The reason-



ing here is that many of these genes probably give rise to cancer of the brain when their expression is perturbed, says Stiles. The development of new mouse models of brain cancer is also seen as a priority.

Part of the award will be used to develop a mentored academic training programme to expose young physician-scientists to contemporary methods and research in neuro-developmental genetics and neuro-oncology. The new training awards will be five-year post-doctoral fellowships. Still in the planning stages, the brain cancer initiative is expected to get under way later this summer. ■

Diane Gershon is Assistant Editor, New Technology, at *Nature Medicine*.